IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.20XX.DOI

CoastXplain: Explainable, Unsupervised Time-Series Modeling of Coastline Changes

CHANDRABALI KARMAKAR[1], ARNAB BHOWMIK[2], ANDRES CAMERO[3], MIHAI DATCU[4]
[1,2,3]German Aerospace Center (DLR), Germany (e-mail: chandrabali.karmakar@dlr.de)
[4]University Politehnica of Bucharest (UPB), Romania (e-mail: mihai.datcu@upb.ro)

Corresponding author: Chandrabali Karmakar (e-mail: chandrabali.karmakar@dlr.de).

ABSTRACT Due to changes in global climate, coastal erosion is a common phenomenon, leading to flooding, habitat loss, property damage, and economic impacts. Large-scale automatic monitoring of coastlines is feasible by processing satellite images with Artificial Intelligence (AI) models; however, it is often constrained by scarce labels and model opacity. We present an explainable, fully unsupervised framework that first converts multi-year Sentinel-2 image series into pixel-wise land–water maps with attached model certainty and then discovers change patterns in the temporal structures of these land-water masks over a specified study period. Throughout the pipeline, we used the Normalized Difference Water Index (NDWI), computed from Sentinel-2 Green (B3) and NIR (B8), as our primary spectral indicator of surface water. We leverage two probabilistic models independently to select the best: (i) a Latent Dirichlet Allocation (LDA) model on a bag-of-visual-words to discover interpretable topics (e.g., foam, water, vegetation) that provide semantic segmentation with transparency, and (ii) a Gaussian Mixture Model (GMM) in pixel feature space to produce per-pixel probabilistic clustering and segmentation. Both models create multiple binary land-water segmentation maps from static images and then a single time-series evolution map from these multi-year land-water segmentation maps. The shoreline was automatically extracted as the boundary land pixels in binary maps with a threshold-based certainty score. We summarize changes with three methods and a score to define the amount of changes: the scene-level water fraction $\widehat{p}_{\mathrm{w}}$, shoreline-normal displacement (SND) along fixed transects, and the low-confidence proportion $\mathcal{U}_{\tau}$ that localizes ambiguity to physically dynamic interfaces. A Hamming score was computed to provide a quick and quantitative overview of the time-series evolution over a 7-year period. The pipeline delivers clear coastlines, interpretable segmentation maps, and uncertainty visualizations to help domain experts in auditing decisions. The result is a label-free, scalable, and explainable workflow in which every decision is accompanied by a confidence score, supporting reliable coastal monitoring and downstream scientific and application-oriented use.

INDEX TERMS Explainable AI (XAI), Coastal Change Monitoring, Sentinel-2 Time-Series, Visualizations.

I. INTRODUCTION

Despite the availability of immense amounts of Earth observation data from various sensors and the open accessibility of many artificial intelligence algorithms, explainable unsupervised classification of Earth observation images remains relatively unexplored. This challenge is twofold: the unavailability of labeled datasets and the need for explainability and human interactability. Most machine learning models are supervised and require expensive and expert-generated labels. Inaccuracies in such labeling can result in model errors. Therefore, leveraging large amounts of unlabeled data is beneficial, although it poses challenges for image annotation and classification.

In contrast, deep learning models have achieved great success in image understanding, natural language processing, and speech recognition. However, they often lack explainability, which is critical in contexts related to global security, such as coastal protection. The European Union's General

Data Protection Regulation, for example, emphasizes the "right to explanation." Adadi et. al list [1] several reasons for requiring explanations: clear decision reasoning, greater control, model improvement, and knowledge discovery. Similarly, researchers have argued that explainability is essential for ensuring scientific validity.

Recent studies have focused on explaining black-box AI models. For instance, Schlegel *et al.* proposed a methodology for evaluating eXplainable Artificíal Intelligence (XAI) methods on time series, while Hohman *et al.* reviewed the role of visual analytics in deep learning. Xi and Panoutsos incorporated fuzzy logic-based layers for interpretability, and Wick *et al.* introduced a "cyclic boosting" algorithm that allows a detailed understanding of predictions [2], [3].However, our contribution in the domain of explainability focuses on intrinsically explainable unsupervised models, as listed below:

OUR CURRENT AND PERVIOUS CONTRIBUTIONS IN XAI

Explainable unsupervised segmentation. A bag-of-visual-words (BoW) + LDA stage yields topic maps that align with physical coastal elements and can be read by experts without labels.We did a similar study with Sentinel-1 image time series for sea-ice change monitoring [4], [5]

Probabilistic land-water segmentation with model certainty. A GMM provides pixel-wise classifications and certainty scores$\gamma_{\max}$ as described in [6].

Time-series structuring and change signatures. We stack monthly segmentations into binary strings per pixel resulting a stack of length 72 denoting 72 time stamps(months of observation) to summarize temporal behavior and compute per-class change scores.We show the semantic maps and evolution maps through an user interface enabling domain-experts' feedbacks. This is a novel contribution of this study to both XAI and human-in-the-loop AI.

Operational summaries. Single-panel triptychs (uncertainty, segmentation, and counts) and shoreline-normal displacement (SND) enable concise reporting.

Our work addresses explainable unsupervised image classification using LDA and GMM to discover latent structures and explain the final outcomes in physical terms. LDA was chosen because it is completely unsupervised, produces interpretable intermediate results, and satisfies key explainability conditions.

A. EXPLAINABILITY XAI

Beyond drawing a crisp coastline, our main goal is that a coastal scientist can understand why the algorithm put the line where it did and how sure it is about that decision. In CoastXplain we therefore treat explainability as a design objective rather than a post-hoc add-on. Concretely, we combine two intrinsically interpretable probabilistic models: a BoW+LDA stage that produces semantic topic maps (foam, open water, sand, vegetation) and a GMM that outputs calibrated per-pixel water probabilities and certainty scores $\gamma_{\max}$. We then turn each pixel's 72-month history into a binary time string and compare change classes using normalized Hamming distance, so that "how much this location changed" is expressed as a simple, percentage-like quantity. In practice, each coastline point carries a small explanation vector saying: what visual pattern we saw, how likely it is to be water, how confident the model is, and how often that decision changed over the record. This is the lens through which we later interpret our results and interact with domain experts in a human-in-the-loop fashion.

B. UNCERTAINTY QUANTIFICATION

A robust environmental monitoring system must generate predictions along with measures of confidence to assess their reliability.

1) Importance of UQ

Uncertainty quantification (UQ) is vital in remote sensing since satellite imagery is subject to noise from atmospheric disturbances, sensor inaccuracies, and natural variability. Quantifying uncertainty helps: **Identify ambiguities:** Highlight regions where the model is less confident. **Enhance decision-making:** Prioritize areas for intervention based on prediction reliability. **Improve model robustness:** Provide feedback for targeted improvements.

2) Types of Uncertainty

Our framework distinguishes between: **Aleatoric Uncertainty:** Inherent data noise that is irreducible. **Epistemic Uncertainty:** Uncertainty from limited data or model imperfections, which can be reduced with additional labeled data.

3) Methods for Quantifying Uncertainty

Our approach leverages probabilistic models—such as unsupervised and semi-supervised Gaussian Mixture Models (GMM)—to produce posterior probabilities, confidence intervals, and entropy-based measures, providing a comprehensive evaluation of uncertainty.

C. HUMAN-IN-THE-LOOP AI

Integrating human expertise into the machine learning pipeline ensures that predictions are both precise and interpretable.

1) Role and Benefits

Human experts: Validate model predictions, especially in regions of high uncertainty. Enhance label quality in semi-supervised learning. Guide model refinement through feedback.

2) Implementation in Our Framework

Our process involves: Generating pseudo-labels with unsupervised models. Expert review and correction to form a reliable ground truth. Iterative refinement using semi-supervised methods. A feedback loop where experts monitor uncertainty metrics.

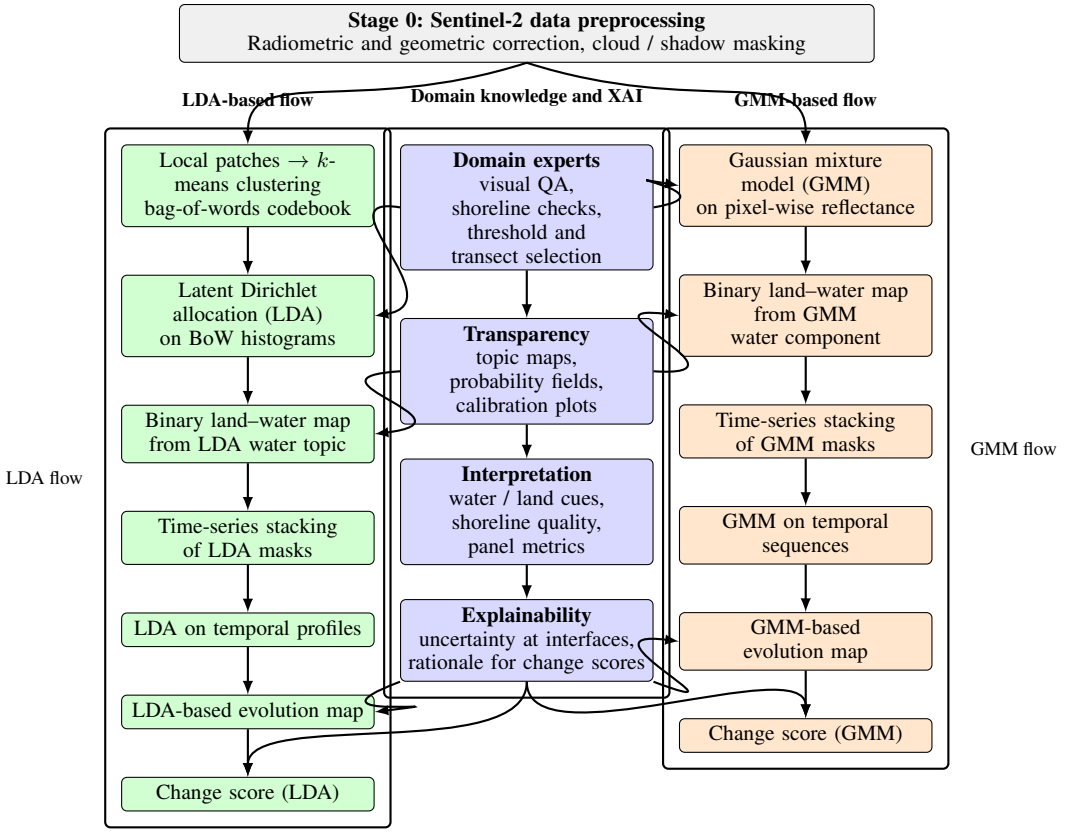


FIGURE 1. High-level overview of the methodology. From the shared Stage 0 Sentinel-2 data preprocessing, two unsupervised flows run in parallel: a bag-of-words + LDA branch (left) and a Gaussian-mixture branch (right). Both produce binary land–water masks, time-series stacks, evolution maps, and per-pixel change scores. Domain experts and explainable-ML elements (transparency, interpretation, explainability) interact with intermediate maps and scores, making each decision point auditable by coastal scientists.

II. STATE OF THE ART

Coastline change detection has become a critical research area due to the accelerating impacts of climate change, sea-level rise, and anthropogenic pressures on coastal environments [7], [9]. The increasing availability of high-resolution, multi-temporal satellite imagery—particularly from missions like Sentinel-2—has enabled the development of advanced machine learning and remote sensing techniques for robust, scalable monitoring of coastal dynamics [9].

A. UNSUPERVISED AND EXPLAINABLE MACHINE LEARNING

Traditional supervised learning approaches for land cover and change detection require extensive labeled datasets, which are often unavailable or costly to obtain for large-scale coastal regions [11]. To address this, unsupervised and semi-supervised models have gained traction. Among these, Latent Dirichlet Allocation (LDA) has emerged as a powerful Bayesian generative model for discovering latent semantic structures in remote sensing data without the need for manual annotation. LDA provides interpretable intermediate representations—such as topic maps and change signatures—that facilitate expert validation and model explainability.

Recent studies have demonstrated the effectiveness of LDA in extracting meaningful land cover classes and detecting change patterns in Sentinel-2 time series. Kim et al. [13] formulated change detection as topic evolution, showing how LDA topics correspond to physical features (e.g., water, sand, vegetation) and their temporal dynamics. These interpretable outputs are crucial for scientific transparency and regulatory compliance, especially under frameworks such as the European Union's "right to explanation" [1].

B. UNCERTAINTY QUANTIFICATION AND PROBABILISTIC MODELS

Uncertainty quantification (UQ) is essential in remote sensing due to inherent noise from atmospheric effects, sensor limitations, and natural variability [15]. Probabilistic models, such as Gaussian Mixture Models (GMMs), are widely used to provide posterior probabilities and confidence intervals for pixel-level classification and pseudo-label generation. By distinguishing between aleatoric (data-inherent) and epistemic (model-related) uncertainty, these methods help prioritize expert review and intervention in ambiguous regions.

C. HUMAN-IN-THE-LOOP AND INTERACTIVE AI

Integrating human expertise into the machine learning pipeline—known as human-in-the-loop (HITL) AI—has

proven highly effective for improving both the accuracy and interpretability of coastal change detection. In this paradigm, unsupervised models generate initial pseudo-labels, which are then reviewed and corrected by domain experts. This iterative feedback loop enables the creation of reliable ground truth datasets and the refinement of semi-supervised models, significantly reducing labeling costs while maintaining high accuracy.

D. ADVANCES IN EXPLAINABLE AI AND VISUAL ANALYTICS

The field of explainable artificial intelligence (XAI) has matured rapidly, with a growing emphasis on transparency and interpretability in environmental monitoring applications [1], [17]. Model-agnostic explanation techniques (e.g., SHAP, LIME) and inherently interpretable algorithms (e.g., cyclic boosting [18], fuzzy logic-based neural networks [19]) are increasingly used to audit and communicate model decisions. Visual analytics tools further support expert understanding by providing interactive exploration of model predictions and uncertainty metrics [2], [3].

E. RECENT TRENDS AND LARGE-SCALE APPLICATIONS

Recent large-scale studies have leveraged deep learning, XAI, and probabilistic models to map and monitor global coastlines with unprecedented detail [7], [9]. Vos et al. [9] present a machine learning pipeline for global shoreline mapping from Sentinel-2, integrating uncertainty estimation and expert validation. Karmakar et al. demonstrate the use of explainable unsupervised models for multi-year coastline change analysis, producing interpretable outputs validated by domain experts.

F. SUMMARY

The convergence of explainable, probabilistic modeling, uncertainty quantification, and human-in-the-loop frameworks represents the state of the art in coastline change detection. These advances enable robust, transparent, and scalable monitoring solutions that are critical for coastal resilience planning and scientific discovery.

III. COMPARISON WITH STATE-OF-THE-ART METHODS

Recent coastal monitoring pipelines—ranging from global satellite-derived shorelines to deep-learning segmentation engines—have substantially advanced automatic shoreline mapping. However, they also exhibit structural limitations in terms of label dependence, uncertainty quantification, and explainability. In this section, we compare CoastXplain against five representative methodologies drawn from the literature: (i) global sandy-beach monitoring from Landsat [7]; (ii) a machine-learning pipeline for global shoreline mapping from Sentinel-2 [9]; (iii) Sentinel-2 time-series land-cover classification using supervised learning [21]; (iv) LDA-based unsupervised change detection in multi-temporal optical imagery [13]; and (v) explainable, LDA-based unsupervised data mining in SAR and its coastal extension [4]–[6]. Together, these methods span traditional SDS, supervised time-series classification, unsupervised topic models, and explainable probabilistic approaches.

A. GLOBAL SANDY-BEACH MONITORING FROM LANDSAT

Luijendijk et al. [7] present one of the earliest genuinely global analyses of sandy shorelines. Their workflow is built on multi-decadal Landsat imagery, from which shorelines are derived using spectral water indices, thresholding and quality control rules. The core idea is to generate a consistent, automated satellite-derived shoreline (SDS) record for 1984–2016, then fit linear trends to quantify rates of erosion and accretion at the global scale. The method is optimized for robustness and scalability: index thresholds and temporal filtering are chosen to handle diverse wave climates, tides and atmospheric conditions with minimal site-specific tuning. This enables an unprecedented bird's-eye view of how sandy beaches have evolved worldwide.

From a methodological standpoint, the Luijendijk pipeline is primarily an index-based, single-epoch approach with strong temporal aggregation. Each individual shoreline is extracted from a single image using a water index (e.g. NDWI) and heuristics to reject outliers. Temporal information enters only when regressing long time series of shoreline positions. Uncertainty is treated implicitly through the density and spread of shoreline observations, and the final product is a single trend estimate per coastal cell. There is no explicit pixel-wise probabilistic classification, and intermediate reasoning (e.g. why a given shoreline is placed at a particular location in a specific epoch) is not exposed to the user.

CoastXplain is complementary and operates at a different point in the design space. Instead of deriving a single polyline per epoch, our framework first constructs a pixel-wise probabilistic land–water segmentation from Sentinel-2 time series and then summarizes dynamics via temporal signatures, change classes and uncertainty metrics. Unlike [7], temporal information is used from the outset: each pixel carries a seven-year binary history, from which we derive change rates, normalized Hamming distances and qualitative change classes. This yields a rich description of how often a location wets and dries, rather than only a long-term shoreline trend. Moreover, every land–water decision is accompanied by a calibrated probability and a certainty measure $\gamma_{\max}$, making it possible to localize ambiguous regions and to gate decisions by confidence. In contrast to the largely heuristic, index-based reasoning in [7], CoastXplain provides interpretable probabilistic layers—topic maps, posterior fields and change signatures—that can be inspected directly by coastal scientists.

B. MACHINE-LEARNING PIPELINE FOR GLOBAL SHORELINE MAPPING

Vos et al. [9] propose a machine-learning pipeline for global shoreline mapping from Sentinel-2 imagery. Building on

the CoastSat concept, their system combines standardized pre-processing (cloud masking, water index computation, tidal filtering) with supervised machine-learning models that distinguish water from land at each acquisition. A key contribution is the integration of big-data cloud computing and carefully designed quality-control routines: candidate shorelines are automatically extracted from water/land masks, filtered based on ancillary information (e.g. tidal level, wave conditions) and combined into global shoreline time series. The method is tuned for operational scalability, delivering near-global coverage with limited human intervention.

The Vos pipeline represents a modern, supervised SDS approach. Its strengths lie in its ability to leverage labeled datasets to learn complex decision boundaries in feature space, and to apply these models at planetary scale. Nevertheless, it inherits several limitations typical of supervised deep or shallow models in EO: (i) the need for representative training labels that capture the full diversity of coastal conditions; (ii) the reliance on black-box classifiers, where internal feature representations and decision rules are difficult to interpret; and (iii) limited explicit uncertainty quantification. While data quality metrics and internal thresholds may be used to flag suspect shoreline segments [9], the method does not provide calibrated per-pixel probabilities or uncertainty maps that could be directly consumed by end users.

CoastXplain takes a different route: it forgoes global labelled training data and instead uses unsupervised probabilistic models (LDA and GMM) to derive per-pixel posteriors from Sentinel-2 time series. Rather than learning a single, global classifier, we allow each scene to be explained by a small number of topics and Gaussian components whose semantics can be inspected by experts. This yields interpretable intermediate objects—topic maps that look like "water", "foam", "wet sand" or "vegetation", and GMM posteriors that behave like calibrated confidence scores. In contrast to the patch- or transect-based aggregation in [9], CoastXplain remains strictly pixel-wise throughout, which preserves fine-scale geomorphic structures such as narrow swash zones and ephemeral channels. Finally, our framework explicitly emphasizes explanation and uncertainty: the products are not only shorelines, but full probability rasters, uncertainty maps, and time-series summaries that answer "where, why and how sure" in a consistent way.

C. SUPERVISED LAND-COVER MAPPING FROM SENTINEL-2 TIME SERIES

García-Haro et al. [21] develop a land-cover mapping pipeline based on Sentinel-2 time series, which is representative of a broader class of supervised, time-series-based classifiers in EO. Their approach extracts spectral and temporal features (e.g. seasonal metrics, vegetation indices) from multi-year Sentinel-2 data and then trains ensemble classifiers to assign each pixel to a land-cover class. By leveraging the full temporal signature rather than single snapshots, the method can separate classes that are spectrally similar in individual images but differ in phenology or persistence. The authors demonstrate that time-series features substantially improve classification performance compared to single-date approaches, especially for complex land-cover mosaics.

Methodologically, the García-Haro pipeline shows how temporal information can be encoded into feature vectors and processed by standard machine-learning algorithms. However, the framework remains fully supervised: it relies on a labelled training set, and the classifier outputs hard labels without exposing the underlying decision surfaces. Uncertainty is generally inferred from confusion matrices or cross-validation statistics, which provide global summaries but no explicit per-pixel uncertainty maps. Moreover, the target is generic land cover; coastlines are only one of many possible interfaces, and the method is not specialized to capture the specific dynamics of intertidal zones or breaker belts.

CoastXplain adopts the idea of exploiting time-series information but in an unsupervised and explainable way. Instead of feeding a vector of temporal features into a supervised classifier, we encode each pixel's seven-year history as a binary time string $b(p)$ and derive qualitative change classes CC-0–CC-3 and Hamming-based dissimilarities. This representation is immediately interpretable to coastal scientists ("hardly ever changes", "changes occasionally", "changes often", "almost always in motion") and can be aggregated over management units. In contrast to [21], where the classifier is a black box and the time series is an internal feature, CoastXplain makes the temporal behaviour itself a first-class explanatory object, tightly coupled to uncertainty summaries and topic semantics.

D. LDA-BASED UNSUPERVISED CHANGE DETECTION IN OPTICAL IMAGERY

Kim et al. [13] propose an unsupervised change-detection method that combines Latent Dirichlet Allocation (LDA) with multitemporal optical imagery. In their framework, local image patches from different dates are quantized into visual words, and LDA is used to discover latent topics that capture common spectral patterns. Change detection is then formulated as the evolution of topic proportions between epochs. Instead of asking "did this pixel change from class A to class B?", the method asks "did the mixture over topics at this location change significantly?", leading to a more nuanced view of gradual or complex changes. The approach is appealing for label-scarce settings: it does not require manual change labels and yields interpretable topic maps and temporal signatures.

E. EXPLAINABLE AI AND INTERPRETABLE MODELS

Explainable artificial intelligence (XAI) has become a major research theme in machine learning and is increasingly relevant for scientific and environmental applications. Adadi and Berrada [1] survey the landscape of XAI techniques, distinguishing between intrinsic interpretability (e.g., models that are transparent by design) and post-hoc explanations (e.g., saliency maps, feature attribution). They also discuss legal and ethical motivations such as the European Union's

"right to explanation", which require that automated systems provide intelligible reasons for their outputs. In Earth observation, such considerations arise whenever ML-derived maps inform policy, planning or regulatory decisions.

Roscher et al. [17] focus on explainable machine learning for scientific discovery. They argue that, in scientific contexts, ML models should not only predict accurately but also help uncover new mechanisms or validate existing theories. The authors propose a taxonomy of XAI methods and illustrate how interpretable models and explanation tools can be used to gain domain insights in fields such as physics and biology. Translating this perspective to coastline monitoring suggests that models should expose physically meaningful intermediate quantities such as change classes, temporal signatures and uncertainty metrics, rather than just producing a binary shoreline.

Visual analytics is a natural companion to XAI. Hohman et al. [2] present visual analytics tools for deep learning that allow experts to explore feature spaces, attribution maps and internal activations through interactive dashboards. Schlegel et al. [3] propose a framework for evaluating XAI methods on time-series models, emphasizing the importance of task-specific evaluation metrics and human studies. These works demonstrate that explanations are most useful when they can be explored interactively and are tightly coupled to the user's tasks.

In addition to post-hoc explanations, there is a line of work on intrinsically interpretable models. Wick and Müller [18] introduce cyclic boosting, a boosting algorithm that yields additive models with transparent effects for each feature. Xi and Panoutsos [19] propose fuzzy-logic-based neural networks that represent learned knowledge in terms of fuzzy rules, thereby combining deep learning with rule-based interpretability. While these models have not yet been specifically applied to coastline monitoring, they illustrate design patterns that could be adapted to coastal contexts: models that remain interpretable by construction while still capturing complex non-linear relationships.

F. HUMAN-IN-THE-LOOP AND INTERACTIVE MACHINE LEARNING

Human-in-the-loop (HITL) and interactive machine learning aim to integrate expert knowledge directly into the learning process. Tuia et al. [24] investigate active learning for remote-sensing image classification, where the algorithm iteratively selects the most informative samples and queries an oracle (e.g., a human expert) for labels. Their results show that carefully designed query strategies can achieve target accuracies with substantially fewer labelled samples than random sampling, particularly in high-dimensional spectral space. Active learning is therefore an attractive strategy for applications where labels are expensive, such as detailed coastal mapping.

Gómez-Chova et al. [23] broaden this view and discuss interactive machine-learning paradigms for Earth observation more generally. They highlight use cases where experts collaboratively refine models through iterative labeling, feedback and visual exploration of intermediate results. Typical system patterns include (i) a machine-learning core that proposes segmentations or class labels, (ii) visualization components that expose uncertainty, disagreement or model attention, and (iii) user interfaces through which experts can correct errors, supply new labels or adjust model hyperparameters. The authors argue that such interactive loops are particularly valuable when users are non-ML-specialists (e.g., environmental scientists or managers) who still need fine-grained control and transparency.

In the context of coastline monitoring, most existing pipelines remain largely offline and unidirectional: models are trained once, applied to large archives, and the resulting shorelines are then visually checked by experts [7], [9]. Few systems allow experts to steer the segmentation process based on spatially localized feedback or to explore uncertainty and model explanations interactively. Recent probabilistic coastal frameworks [6] begin to close this gap by combining unsupervised models with uncertainty maps that can guide expert intervention (for example, by highlighting ambiguous shoreline segments), but they still stop short of fully integrated interactive dashboards as envisioned in the interactive ML literature [2], [23].

G. SUMMARY AND OPEN GAPS

The state of the art in coastline monitoring thus encompasses a rich spectrum of methods:

1) **Traditional pipelines** based on spectral indices, thresholding and heuristics remain widely used and can scale to global archives [7], but handle uncertainty and complex coastal morphology only indirectly.
2) **Supervised machine learning and deep learning** achieve high mapping accuracy and leverage spatial and temporal context [11], [21], yet rely on large labelled datasets and often behave as black boxes without calibrated uncertainty [22].
3) **Unsupervised topic models and generative approaches** such as LDA capture semantic structure and temporal dynamics without labelled data and produce interpretable latent representations [4], [5], [13], but have so far been applied to limited coastal case studies and rarely integrated into operational coastal products [6].
4) **Uncertainty quantification frameworks** in remote sensing and deep learning provide conceptual and algorithmic tools to represent predictive uncertainty [15], [22], but explicit uncertainty maps are still the exception rather than the rule in shoreline products [11].
5) **Explainable and interactive ML** offer visual analytics, interpretable models and HITL workflows that can increase trust and scientific value [1]–[3], [17]–[19], [23], [24], yet these ideas are only beginning to be specialized to coastline monitoring.

Overall, existing methodologies tend to optimize one axis at a time: either accuracy and automation (deep-learning

and SDS pipelines), or interpretability (topic models, interpretable boosting, fuzzy-rule networks), or explicit uncertainty, or human-in-the-loop interaction. There is still a lack of integrated frameworks that jointly provide:

1) largely unsupervised learning from long optical time series,
2) per-pixel and segment-wise uncertainty quantification,
3) explainable intermediate representations such as qualitative change classes and temporal signatures, and
4) a human-in-the-loop interface that allows experts to explore, validate and refine the model outputs.

The CoastXplain framework proposed in this work is designed to address precisely this gap: it combines unsupervised, probabilistic modelling of Sentinel–2 time series with explicit uncertainty metrics and an interactive, explainable interface for coastal managers. Rather than competing solely on segmentation accuracy with deep CNNs, CoastXplain aims to complement existing state-of-the-art methods by providing transparent, uncertainty-aware summaries of coastal dynamics that are better aligned with decision-making and scientific interpretation. CoastXplain builds on the LDA building block but extends it in several directions tailored to coastal dynamics. First, we use LDA not only on spatial bag-of-words histograms but also on temporal profiles, yielding evolution maps and qualitative change classes that summarize multi-year wetting–drying behaviour. Second, we treat LDA as part of a broader probabilistic ensemble: its soft water scores are fused with GMM posterior probabilities, and both are calibrated so that thresholds have a consistent meaning across sites and years. Third, the LDA outputs are embedded into an explanation narrative: topic maps, soft water scores, $\gamma_{\max}$ and Hamming-based change signatures are jointly displayed in GIS-ready panels, making it clear where coastal interfaces lie, which mixtures of topics support them and how often they shift. In this way, CoastXplain can be seen as a coastal-specialized, uncertainty-aware and human-facing extension of the generic LDA-based change-detection paradigm in [13].

H. EXPLAINABLE UNSUPERVISED COASTAL MODELS AND XAI FRAMEWORKS

A closely related line of work is the use of LDA as an explainable unsupervised model for SAR and coastal imagery. Karmakar et al. [4], [5] apply LDA to SAR data to discover semantic structures such as sea-ice regimes and land-cover types. Their "feature-free explainable data mining" framework treats LDA topics as semantic layers: each topic corresponds to visually coherent patterns that experts can interpret (e.g. open water, different ice types). This approach demonstrates that topic models can act as intrinsic explanation mechanisms in remote sensing: rather than extracting features and then training a black-box classifier, the generative model itself provides human-readable latent variables.

In subsequent work, Karmakar et al. [6] extend unsupervised probabilistic models to coastline extraction, introducing uncertainty-aware unsupervised learning to draw coastlines. Their coastal framework combines LDA and Gaussian mixture models to derive land–water segmentation and shoreline positions, while explicitly reporting model uncertainty. This is already close in spirit to CoastXplain: both frameworks use unsupervised probabilistic models, aim for interpretability via topic maps and posterior probabilities, and acknowledge the importance of uncertainty quantification in coastal monitoring.

CoastXplain can be viewed as a systematization and further generalization of these ideas. Compared to the SAR-focused applications in [4], [5], CoastXplain is explicitly designed for long Sentinel-2 optical time series, multi-year change signatures and qualitative change classes. It integrates LDA and GMM into a single pipeline that produces not only segmentations and shorelines, but also calibrated probabilities, per-pixel certainty, temporal Hamming distances, and human-readable change classes CC-0–CC-3. In contrast to many XAI methods that are post-hoc add-ons to black-box networks [1]–[3], [17], CoastXplain is intrinsically explainable: its intermediate representations are designed from the outset to be physically meaningful and auditable. The framework also aligns with broader calls for uncertainty-aware and human-in-the-loop EO systems [15], [22], [23], by providing uncertainty maps and compact indicators (e.g. p_{bw}, SND, U_τ) that can guide expert attention and operational decision-making.

I. OVERALL POSITIONING

Across these representative methodologies, a common pattern emerges. Global index-based pipelines excel at coverage but offer limited uncertainty and interpretability [7]; supervised machine-learning and deep-learning approaches achieve high accuracy given labels but behave as black boxes and often lack explicit uncertainty [9], [21], [22]; unsupervised topic models and probabilistic coastal frameworks provide interpretable semantics but have so far been applied to specific sensors or regions [4]–[6], [13]; and general XAI frameworks emphasize explanation but rarely specialize to the peculiarities of coastal time series and intertidal dynamics [1], [17], [23].

CoastXplain is designed to occupy the intersection of these strands: it is unsupervised, probabilistic, intrinsically explainable, explicitly uncertainty-aware and tailored to multi-year Sentinel-2 coastline monitoring. Rather than competing purely on pixel-wise accuracy with global deep networks, it aims to deliver a transparent, auditable and operationally usable representation of coastal dynamics, in which every map, time series and shoreline is accompanied by an explicit story of "what changed, where, why, and how sure".

IV. METHODOLOGY

Our aim is to turn multi-year Sentinel-2 time series into (i) a robust, pixel-wise land–water segmentation and (ii) uncertainty summaries that show where results are fragile and why. We explicitly experiment with two unsupervised, probabilis-

tic models—Latent Dirichlet Allocation (LDA) and Gaussian Mixture Models (GMM)—to generate per-pixel posteriors; we interpret these probabilities as model certainty and report them alongside the binary mask. Figure **??** provides the bird's-eye view. Below, we walk through the full pipeline in the way a practitioner would build and debug it, highlighting design choices, failure modes, and safeguards.

A. DATASET DESCRIPTION

Time-series imagery was acquired monthly from January 2018 to September 2023, covering 4,694 cloudfree locations along the North Sea and Baltic Sea coastlines (each 5 km × 5 km). Locations were selected using farthest point sampling for representative coverage. The imagery was subdivided into scenes of 1.28 km × 1.28 km.

Data layout and notation.

For each of $L{=}8$ coastal locations, we acquire a monthly Sentinel-2 series of $T{=}72$ months. Each scene is partitioned into a $H{\times}W$ grid ($H{=}W{=}480$). Let $p \in \{1,\dots,HW\}$ index pixels, $t \in \{1,\dots,T\}$ index months, and $c \in \{B2, B3, B4, B8\}$ index bands. Denote the top-of-atmosphere reflectance by $I_c^{(t)}(p)$, and the stacked multiband image at month t by $\mathbf{I}^{(t)} \in \mathbb{R}^{H\times W\times 4}$.

B. STAGE 0: ANALYSIS-READY DATA (ARD) WITH COASTAL PITFALLS IN MIND

Bands and co-registration. We work with B2/B3/B4/B8. All scenes are co-registered to a common UTM grid via feature-based alignment (keypoints on stable inland areas) followed by a sub-pixel Lucas–Kanade refinement on image pyramids. We estimate the residual misalignment as

$$\varepsilon_{\text{reg}} = \frac{1}{|\Omega_{\text{stable}}|} \sum_{p\in\Omega_{\text{stable}}} \left\| \nabla \mathbf{I}^{(t)}(p) - \nabla \mathbf{I}^{(t_0)}(p) \right\|_2, \quad (1)$$

and require $\varepsilon_{\text{reg}} < \tau_{\text{reg}}$ (empirically $\tau_{\text{reg}}{\approx}0.1$ in normalized units); otherwise the scene is flagged.

Radiometry. To temper illumination/seasonality, we standardize each band using robust statistics:

$$\tilde{I}_c^{(t)}(p) = \frac{I_c^{(t)}(p) - \text{median}(I_c^{(t)})}{1.4826\ \text{MAD}(I_c^{(t)})}, \quad c \in \{B2, B3, B4, B8\}. \quad (2)$$

This transformation is applied per-scene, per-band, and preserves rank order while bounding outlier influence.

Water index (NDWI). As our main spectral indicator of surface water we compute, for each month t, the Normalized Difference Water Index

$$\text{NDWI}^{(t)}(p) = \frac{I_{B3}^{(t)}(p) - I_{B8}^{(t)}(p)}{I_{B3}^{(t)}(p) + I_{B8}^{(t)}(p)}, \quad (3)$$

using the Green (B3) and NIR (B8) bands. NDWI is used for quick visual QA, for coastline inspection, and as auxiliary evidence in later stages of the pipeline (e.g. topic scoring).

Quality masks ("blue–NIR veto"). Provider cloud masks are conservative near bright foam; we therefore add an auxiliary veto that flags (not removes) pixels:

$$\mathcal{Q}^{(t)}(p) = \mathbf{1}\Big\{ \tilde{I}_{B2}^{(t)}(p) > \tau_{B2} \ \wedge\ \tilde{I}_{B8}^{(t)}(p) < \tau_{B8} \ \wedge\ \max_{q\in\mathcal{N}_3(p)} \tilde{I}_{B2}^{(t)}(q) > \tau \quad (4)$$

with $\tau_{B2}{\approx}2.0$, $\tau_{B8}{\approx}{-}0.5$ and $\mathcal{N}_3(p)$ the $3{\times}3$ neighborhood. These pixels are carried forward with reduced confidence rather than being dropped.

Tides. We do not tide-normalize scenes; instead, we treat waterline mobility as part of the phenomenon and account for it in the uncertainty summaries and displacement metrics.

C. STAGE 1: LOCAL EVIDENCE AS SMALL, OVERLAPPING MICROPATCHES

Coastlines are texture-rich at 10 m scale (foam, ripples, wet sand). We therefore describe each location by overlapping micropatches P_j of size $s \times s$ (default $4{\times}4$) with stride $r \in \{2, 4\}$. For each P_j we compute a compact descriptor

$$\mathbf{f}_j = \Big[\mu(\tilde{\mathbf{I}}_{P_j}),\ \sigma(\tilde{\mathbf{I}}_{P_j}),\ \nabla_x\tilde{\mathbf{I}}_{P_j},\ \nabla_y\tilde{\mathbf{I}}_{P_j}\Big] \in \mathbb{R}^D, \quad (5)$$

where μ and σ are per-band means and standard deviations, and ∇_x, ∇_y are Sobel gradients aggregated by average magnitude. This balances spectral separation (mean/variance) and boundary sharpness (gradients). It is intentionally simple: stable, fast, and interpretable.

D. STAGE 2: BAG-OF-VISUAL-WORDS (BOW) TO COMPRESS VARIABILITY

We learn a codebook $\mathcal{C} = \{\mathbf{c}_1,\dots,\mathbf{c}_K\}$ by k-means over a stratified sample of descriptors (shoreline and inland). Each descriptor is quantized

$$w_j = \arg\min_k \|\mathbf{f}_j - \mathbf{c}_k\|_2 \in \{1,\dots,K\}, \quad (6)$$

and for each document (tile or patch-of-patches) we accumulate a BoW histogram $\mathbf{x} \in \mathbb{N}^K$. BoW compresses local variability while preserving the "ingredients list" of visual patterns—perfect feedstock for topic discovery. We select K by minimizing the within-cluster dispersion plateau (knee in $\sum_j \min_k \|\mathbf{f}_j - \mathbf{c}_k\|_2^2$) and by downstream LDA coherence (below).

E. STAGE 3: UNSUPERVISED TOPIC DISCOVERY (LDA) WITH INTERPRETABLE, PROBABILISTIC OUTPUTS

We model $\mathbf{x}$ with LDA to uncover recurring visual topics (e.g., bright foam + blue water + low NIR). For document d with words $\{w_{dn}\}_{n=1}^{N_d}$,

$$\theta_d \sim \text{Dir}(\alpha), \qquad \phi_t \sim \text{Dir}(\beta), \quad (7)$$

$$z_{dn} \mid \theta_d \sim \text{Cat}(\theta_d), \quad (8)$$

$$w_{dn} \mid z_{dn}, \Phi \sim \text{Cat}(\phi_{z_{dn}}), \quad (9)$$

TABLE 1. Sentinel-2 MSI spectral bands used in this study. Band limits are given as approximate spectral range for Sentinel-2A.

Band	Name	$\lambda_{\min}$ [nm]	$\lambda_{\max}$ [nm]	Primary sensitivity / typical use
B1	Coastal aerosol	432	453	Aerosols and atmospheric correction over coastal and hazy scenes.
B2	Blue	459	525	Bathymetry in clear water, soil–vegetation discrimination, general true-colour blue channel.
B3	Green	542	578	Vegetation vigor, water turbidity, and general true-colour green channel.
B4	Red	649	680	Chlorophyll absorption, vegetation type mapping, built-up areas; true-colour red channel.
B5	Vegetation red edge 1	697	712	First red-edge band; sensitive to chlorophyll content and early vegetation stress.
B6	Vegetation red edge 2	733	748	Second red-edge band; canopy structure and leaf area index.
B7	Vegetation red edge 3	773	793	Third red-edge band; advanced vegetation status and species separation.
B8	NIR (broad)	780	886	Near-infrared; biomass and canopy vigor, shoreline mapping, NDVI-type indices.
B8A	Narrow NIR	854	875	Narrow NIR; detailed vegetation and canopy structure, chlorophyll content.
B9	Water vapour	935	955	Water-vapour absorption; atmospheric correction over land and water.
B10	SWIR – Cirrus	1358	1389	Detection of thin, high cirrus clouds.
B11	SWIR 1	1568	1659	Vegetation and soil moisture, snow/cloud discrimination, burned-area mapping.
B12	SWIR 2	2115	2290	Geological features, snow/ice and cloud screening, burned-area and smoke plumes.

with topic proportions θ_d and topic–word distributions $\Phi = \{\phi_t\}_{t=1}^{T_{\text{topics}}}$. The joint and evidence are

$$p(\mathbf{w}, \mathbf{z}, \Theta, \Phi \mid \alpha, \beta) = \prod_t \mathrm{Dir}(\phi_t \mid \beta) \prod_d \mathrm{Dir}(\theta_d \mid \alpha) \prod_n \theta_{d,z_{dn}} \phi_{z_{dn},w_{dn}}, \tag{10}$$

$$\log p(\mathbf{w} \mid \alpha, \beta) \geq \mathcal{L}(\lambda, \gamma, \rho) \quad \text{(ELBO)}. \tag{11}$$

We fit via mini-batch variational Bayes; we monitor the ELBO and stop when the relative change falls below 10^{-4}. Topic quality is assessed by topic coherence (NPMI):

$$\mathrm{C}_{\text{npmi}}(t) = \frac{1}{\binom{M}{2}} \sum_{i<j} \frac{\log \frac{P(w_i,w_j)}{P(w_i)P(w_j)}}{-\log P(w_i, w_j)}, \tag{12}$$

computed on the top-M words of topic t using document co-occurrences. We choose T_{topics} that maximizes the mean coherence subject to stability across seeds.

From the fitted model we construct topic maps $\mathbf{T}(p) \in \Delta^{T_{\text{topics}}-1}$ by averaging responsibilities of overlapping micropatches—these are smooth, interpretable layers that already look like "waterish" or "sandish" fields to a human expert.

Human-readable semantics without labels.

We select topic sets $\mathcal{T}_{\text{w}}$ (water-like) and $\mathcal{T}_{\text{l}}$ (land-like) by inspecting ϕ_t (which BoW words the topic uses) and cross-checking spatial context. We also score topics by a weak prior using NDWI-like evidence:

$$\pi_t^{\text{water}} \propto \sum_k \phi_{t,k} \underbrace{\mathbb{E}_{\mathbf{f}|k}\big[\tilde{I}_{B3} - \tilde{I}_{B8}\big]}_{\text{water index proxy}}, \tag{13}$$

and prefer topics with π_t^{water} above a percentile threshold to populate $\mathcal{T}_{\text{w}}$. The soft water score is then

$$s_{\text{w}}(p) = \sum_{t \in \mathcal{T}_{\text{w}}} T_t(p), \qquad s_{\text{l}}(p) = 1 - s_{\text{w}}(p). \tag{14}$$

LDA is purely unsupervised; its per-pixel scores are probabilistic, and we report them as one view of model certainty.

F. STAGE 4: PROBABILISTIC REFINEMENT (FEATURE-SPACE DENSITY + CALIBRATION)

Topics capture which ingredients appear; we now refine how confidently a pixel belongs to water/land in feature space. We fit an unsupervised Gaussian mixture on pixel-wise features $\mathbf{x}(p)$ (band values + local stats):

$$p(\mathbf{x}(p)) = \sum_{k=1}^{K^\star} \pi_k \mathcal{N}(\mathbf{x}(p) \mid \mu_k, \Sigma_k). \tag{15}$$

We choose $K^\star$ by the Bayesian Information Criterion (BIC). **Unsupervised, probabilistic output:** the GMM is fit without labels. Its posteriors yield a water probability

$$\tilde{p}_{\text{w}}(p) = p(\text{water} \mid \mathbf{x}(p)) \tag{16}$$

and a certainty score

$$\gamma_{\max}(p) = \max_k p(k \mid \mathbf{x}(p)), \tag{17}$$

which we report throughout as model certainty.

Because raw posteriors can be over-confident, we calibrate them with a single temperature T by minimizing the Brier score

$$\tilde{p}_{\text{w}}^{\text{cal}} = \sigma\left(\frac{\log \frac{\tilde{p}_{\text{w}}}{1-\tilde{p}_{\text{w}}}}{T}\right), \quad T^\star = \arg\min_T \frac{1}{|\mathcal{V}|} \sum_{p \in \mathcal{V}} \big(\tilde{p}_{\text{w}}^{\text{cal}}(p) - y(p)\big)^2, \tag{18}$$

on a validation split $\mathcal{V}$ with weak labels (e.g., from stable inland/water masks) or via reliability diagrams using temporal self-consistency. As a secondary check we compute the Expected Calibration Error (ECE):

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left|\mathrm{acc}(B_m) - \mathrm{conf}(B_m)\right|. \tag{19}$$

Lightweight spatial consistency.

We add a first-order total-variation prior to discourage salt-and-pepper noise without blurring the waterline:

$$\min_{y\in\{0,1\}^{HW}} \sum_{p} \ell\big(y(p), \tilde{p}_{\mathrm{w}}^{\mathrm{cal}}(p)\big) + \lambda \sum_{(p,q)\in\mathcal{N}} |y(p) - y(q)|. \tag{20}$$

We keep λ small so that the boundary remains faithful to the data. The loss ℓ is cross-entropy with class-balancing from the current scene histogram.

G. STAGE 5: TRANSPARENT FUSION AND BINARY DECISION

Having two unsupervised, probabilistic views—semantic from LDA and feature-space from GMM—we produce a pixel-wise land–water segmentation by fusing their calibrated probabilities:

$$\hat{y}(p) = \mathbf{1}\Big\{\alpha\, s_{\mathrm{w}}(p) + (1-\alpha)\, \tilde{p}_{\mathrm{w}}^{\mathrm{cal}}(p) \geq \tau_b\Big\}, \tag{21}$$

with $\alpha \in [0,1]$ and a conservative threshold $\tau_b \in [0.5, 0.6]$. Equation (21) is the Bayes decision under 0–1 loss with prior odds encoded by α and calibrated likelihoods from the GMM. In the log-odds domain the fusion is linear:

$$\mathrm{logit}\ p_{\mathrm{fused}}(p) = \omega_1\ \mathrm{logit}\ s_{\mathrm{w}}(p) + \omega_2\ \mathrm{logit}\ \tilde{p}_{\mathrm{w}}^{\mathrm{cal}}(p),\ \ \omega_{1,2} \geq 0, \tag{22}$$

with $\omega_{1,2}$ reparameterizing α. Small isolated components ($< \eta$ pixels) are removed; holes are closed with a single binary opening+closing cycle $\mathcal{M}_{\mathrm{open}} \circ \mathcal{M}_{\mathrm{close}}$ with a 3×3 structuring element. We report per-pixel certainty as $\gamma_{\max}(p)$ and keep it visible in all figures.

H. STAGE 6: SHORELINE EXTRACTION AND SUB-PIXEL POSITIONING

The shoreline is extracted as the 0.5-probability iso-contour of $\tilde{p}_{\mathrm{w}}^{\mathrm{cal}}$ via marching squares with bilinear interpolation, yielding sub-pixel vertex positions $\bar{\mathbf{x}}_{\mathrm{shore}}(t) \in \mathbb{R}^2$. Numerical stability is improved by clamping $\tilde{p}_{\mathrm{w}}^{\mathrm{cal}} \in [\epsilon, 1-\epsilon]$ with $\epsilon = 10^{-3}$. For displacement we use cross-shore transects with unit normals $\mathbf{n}$ and compute

$$\mathrm{SND} = \Big\langle \mathbf{n},\ \bar{\mathbf{x}}_{\mathrm{shore}}(t_b) - \bar{\mathbf{x}}_{\mathrm{shore}}(t_a) \Big\rangle, \tag{23}$$

where $\bar{\mathbf{x}}_{\mathrm{shore}}(t)$ is the along-transect shoreline intersection at time t. Positive SND means seaward motion.

Uncertainty on SND.

Let $\sigma_{\parallel}^2$ be the variance of the along-normal position induced by probability noise $\delta\tilde{p}$ and local gradient $g = \|\nabla \tilde{p}_{\mathrm{w}}^{\mathrm{cal}}\|$. A first-order propagation gives

$$\sigma_{\parallel}^2 \approx \frac{\mathrm{Var}[\delta\tilde{p}]}{g^2}, \qquad \widehat{\mathrm{SE}}(\mathrm{SND}) = \sqrt{\sigma_{\parallel,t_a}^2 + \sigma_{\parallel,t_b}^2}. \tag{24}$$

We visualize this as thin confidence ribbons on transect plots.

I. STAGE 7: SINGLE-PANEL SUMMARIES THAT CARRY BOTH RESULT AND DOUBT

Each single panel shows (left) an uncertainty summary derived from $\gamma_{\max}$, (center) the binary segmentation $\hat{y}$, and (right) the land–water counts. From each panel we extract:

$$\widehat{p}_{\mathrm{w}} = \frac{\#\{\hat{y}=1\}}{H\,W}, \qquad \mathcal{U}_\tau = \frac{1}{H\,W} \sum_{p} \mathbf{1}\{\gamma_{\max}(p) < \tau\}, \tag{25}$$

with τ typically 0.7. Over time, trends in $\widehat{p}_{\mathrm{w}}$ describe wet-surface occupancy; ribbons of elevated $\mathcal{U}_\tau$ cling to physically dynamic interfaces (swash, inlets, turbid plumes). Inland areas should remain both stable and confident; if they do not, we treat the scene as radiometrically suspect.

J. TEMPORAL STACKING AND CHANGE SIGNATURES

Once per-epoch masks are available, each pixel carries a binary time string

$$\mathbf{b}(p) = [b_1(p), \ldots, b_T(p)]^\top \in \{0,1\}^T, \qquad b_t(p) = \hat{y}^{(t)}(p). \tag{26}$$

We summarize its behaviour via

$$\mathrm{rate}(p) = \frac{1}{T-1} \sum_{t=2}^{T} \mathbf{1}\{b_t(p) \neq b_{t-1}(p)\}, \tag{27}$$

$$\mathcal{H}(p) = -\frac{1}{T} \sum_{t=1}^{T} \big[b_t \log b_t + (1-b_t)\log(1-b_t)\big], \tag{28}$$

$$d_{\mathrm{Ham}}(\mathbf{b}, \mathbf{b}') = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{b_t \neq b'_t\}. \tag{29}$$

We compute change signatures by averaging $\mathbf{b}(p)$ over class masks and months, and summarize pairwise dissimilarities among qualitative change classes (CC-i) using d_{Ham}.

K. DESIGN CHOICES, WITH THE "WHY"

Micropatches over pixels. Patches stabilize texture cues (foam vs. wet sand) and reduce false toggles at the waterline. **BoW + LDA.** BoW is fast and robust; LDA turns it into interpretable topic maps that experts can discuss ("this topic is foam"). **Calibration.** A single temperature parameter markedly improves the trustworthiness of certainty maps—vital when results are used operationally. **Gentle regularization.** We keep the boundary sharp and let uncertainty—not smoothing—communicate ambiguity. **Sub-pixel shoreline.** The 0.5-contour of a calibrated probability field provides consistent geometry and a principled definition of "the line".

L. ALGORITHMS (FOR REPRODUCIBILITY)

Stage-1–3 are summarized in Algorithm 1; Stage-4–6 in Algorithm 2. Hyperparameters shown below are the defaults we use unless a site requires adjustment.

Stage A: Encoding → Topics.

Algorithm 1 Local Encoding and LDA Topic Discovery

1: **Input:** Bands $B2, B3, B4, B8$, patch size s, stride r, codebook size K, topics T_{topics}
2: Robust-normalize bands (Eq. 2); build micropatches $\{P_j\}$ and descriptors $\{\mathbf{f}_j\}$
3: Fit k-means $\Rightarrow$ codebook $\mathcal{C}$; quantize $\mathbf{f}_j \mapsto w_j$
4: Aggregate BoW $\mathbf{x}$ per document; fit LDA (mini-batch VB, ELBO tol 10^{-4})
5: Compute topic coherence; adjust K and T_{topics} if needed
6: Construct topic maps $\mathbf{T}(p)$ from overlapping responsibilities
7: **Output:** $\mathbf{T}(p)$, Φ, Θ

Stage B: Fusion → Binary + Certainty.

Algorithm 2 Probabilistic Refinement, Calibration, Fusion, and Shoreline

1: Identify $\mathcal{T}_{\text{w}}$ by inspecting Φ and spatial context; compute $s_{\text{w}}(p)$
2: Fit GMM on $\mathbf{x}(p)$; get $\tilde{p}_{\text{w}}(p)$ and $\gamma_{\max}(p)$ **(BIC for $K^\star$)**
3: Calibrate with temperature T (minimize Brier score, check ECE)
4: Fuse via Eq. (21) with (α, τ_b); apply small TV regularization
5: Morphological clean-up; extract shoreline at $\tilde{p}_{\text{w}}^{\text{cal}} = 0.5$
6: Compute SND and its standard error along transects
7: Build panel metrics: $\widehat{p}_{\text{w}}$, $\mathcal{U}_\tau$; store $\gamma_{\max}$ map
8: **Output:** $\hat{y}$, shoreline polyline, $\gamma_{\max}$, panel metrics

M. COMPLEXITY, PARAMETERS, AND PRACTICAL DEFAULTS

Runtime. Encoding and BoW are $\mathcal{O}(HW)$ and trivially parallel. LDA scales with T_{topics}, codebook K, and iterations; mini-batches keep wall-time modest. GMM scales with components $K^\star$ and feature dimension D; diagonal covariances suffice for our spectra. **Defaults.** s=4, r=2, $K \in [128, 256]$, $T_{\text{topics}} \in [8, 12]$, α=0.4, τ_b=0.55, τ=0.7, η=0.001 · HW, small TV weight λ. **When things go wrong.** (i) Over-bright foam misread as land: increase α (more topic emphasis) or raise τ_b. (ii) Backshore shadows misread as water: rely more on NIR by adding a per-band weight in $\mathbf{x}(p)$. (iii) Widespread low confidence inland: treat as radiometric mismatch; re-normalize with scene-wise robust stats or exclude the scene from temporal trend calculations.

N. WHAT THE NUMBERS MEAN TO A DOMAIN EXPERT

$\widehat{p}_{\text{w}}$ rises when wet-surface occupancy increases (tide, surge, erosion); **SND** provides direction and magnitude along physically meaningful transects; $\mathcal{U}_\tau$ localizes ambiguity to where physics is messy (swash, inlet jets) rather than spreading it uniformly. Read together, these three form a concise, trustworthy story about what changed and how confident we are.

O. COASTLINE SHIFT COMPUTATION

To assess the amount of coastline change between two distinct time instants, t_i and t_j, we first identify the coastline positions at these times. Let $\text{pos}(x_{t_i})$ denote the position (or coordinate) of a particular point on the coastline at time t_i, and $\text{pos}(x_{t_j})$ the position of the corresponding coastline point at time t_j.

$$\Delta = \left|\text{pos}(x_{t_i}) - \text{pos}(x_{t_j})\right| \tag{30}$$

Here, Δ is the absolute difference between these two positions, representing the magnitude of the coastline change (i.e., how far the coastline has shifted) between t_i and t_j.

P. BINARY LAND–WATER MAPPING AND UNCERTAINTY SUMMARIES

Each analysis panel is a single composite image containing: (left) an uncertainty summary derived from per-pixel posterior membership $\gamma_{\max}$; (centre) a binary land–water segmentation; and (right) a bar chart of land–water pixel counts. From each panel we extract:

$$\begin{aligned} \widehat{p}_{\text{w}} &= \frac{\#\{\text{water pixels}\}}{H\,W}, \\ \text{SND} &= \langle \mathbf{n},\, \bar{\mathbf{x}}_{\text{shore}}(t_b) - \bar{\mathbf{x}}_{\text{shore}}(t_a) \rangle\,, \\ \mathcal{U}_\tau &= \frac{1}{H\,W} \sum_{i=1}^{H\,W} \mathbf{1}\{\gamma_{\max}(i) < \tau\} \end{aligned} \tag{31}$$

where $\widehat{p}_{\text{w}}$ is the water fraction, SND is shoreline-normal displacement (positive = seaward, negative = landward) along the unit normal $\mathbf{n}$, and $\mathcal{U}_\tau$ is the proportion of low-confidence pixels under threshold τ.

V. RESULTS

The following Sentinel-2 bands have been found to be particularly effective for coastal analysis:

1) **B2 (Blue):** Enhances water detection.
2) **B3 (Green):** Improves the contrast between water and land.
3) **B4 (Red):** Crucial for distinguishing land-water boundaries.
4) **B8 (Near Infrared):** Useful for separating water from vegetation and built-up areas.

we compute the Normalized Difference Water Index (NDWI) from B3 and B8 and use it as our main water index throughout the experiments.

A. BINARY CLASS

Each panel is a single image composed of: (left) uncertainty, (centre) segmentation (land vs. water), and (right) land–water pixel-count bar chart. We report change via $(\widehat{p}_{\text{w}}, \text{SND}, \mathcal{U}_\tau)$.

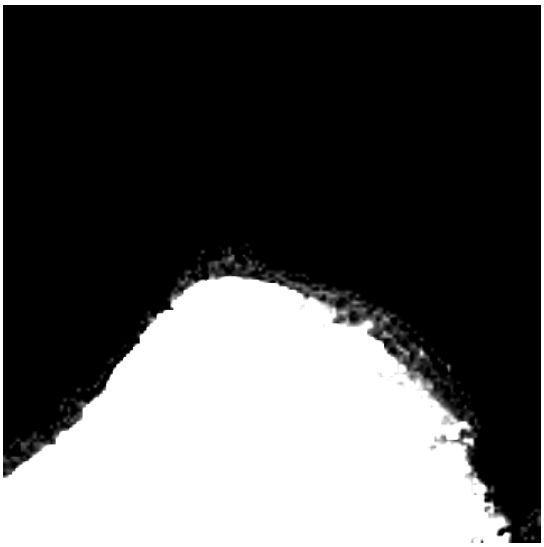


FIGURE 2. Coastal image with DEM

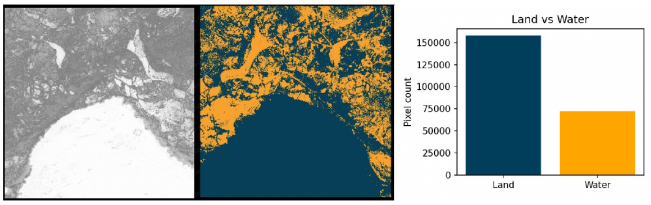


FIGURE 3. Panel A: Uncertainty (left), segmentation (centre), and land–water counts (right).

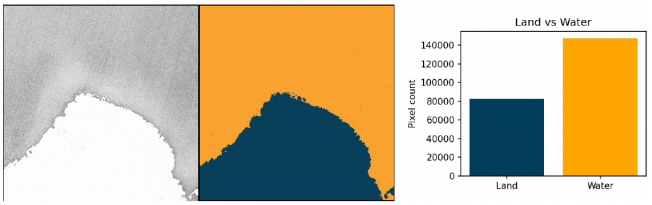


FIGURE 4. Panel B: Uncertainty, segmentation, and land–water counts.

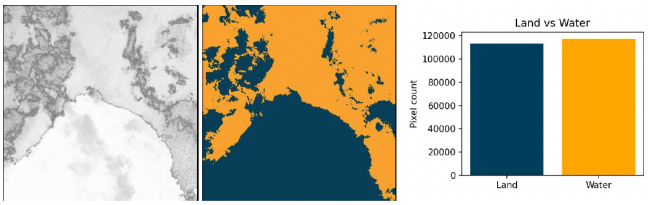


FIGURE 5. Panel C: Uncertainty, segmentation, and land–water counts.

B. CROSS-SITE GENERALIZATION ACROSS NORTH SEA AND BALTIC COASTLINES

Beyond the illustrative panels in Figs. 3–6, we evaluated CoastXplain across all 4,694 locations in our North Sea and Baltic Sea dataset. Despite substantial variations in tidal range, sediment type, and anthropogenic pressure, the pipeline behaved in a consistent, physically plausible way.

First, the fraction of low-confidence pixels $\mathcal{U}_\tau$ remains systematically concentrated near dynamic interfaces such as swash zones, inlets, and turbid plumes. Inland and offshore regions exhibit low $\mathcal{U}_\tau$ except in scenes flagged as radiometrically suspect (e.g., thin cirrus, partial snow). This spatial pattern is robust even when moving between macrotidal sandy coasts and microtidal, engineered shorelines, suggesting that the certainty maps are driven primarily by mixtures of spectra and geometry rather than site-specific tuning.

Second, topic semantics transfer well between regions. Topics identified as "water-like" in one coastal sector (low NIR, high blue/green, smooth textures) retain their interpretation when applied to other sites. Similarly, "sand/foreshore" and "vegetation/backshore" topics remain visually recognizable, allowing experts to reuse the same semantic vocabulary when validating different locations. This cross-site stability greatly eases human-in-the-loop inspection: once topic semantics are understood for a few representative scenes, they remain meaningful across the wider coastline.

Finally, Hamming-based change signatures aggregate smoothly over larger units such as coastal cells or management polygons. Regions known to be morphologically stable (e.g., heavily engineered harbour fronts) cluster near low transition rates and low Hamming distances, whereas open, sediment-rich stretches display higher transition rates and clear separation between stable (CC-0) and active (CC-1–CC-3) classes, as summarized in Table 2. This behaviour indicates that the proposed change metric is not tied to a particular site, but scales to heterogeneous coastlines while preserving a meaningful ordering of "more" versus "less" dynamic sectors.

C. PANEL A

Water share is lower than land, consistent with exposed intertidal/supratidal zones. Uncertainty filaments follow swash and shallow subtidal arcs, indicating mixed spectra and adjacency at boundaries.

D. PANEL B

The water fraction rises relative to Panel A, indicating reduced exposed intertidal area and landward displacement at several transects. Inland confidence remains high.

E. PANEL C

Relative to 2018, several segments shift landward; the shoreline arc smooths, consistent with cumulative erosion. Uncertainty increases only in hydrodynamically active zones.

F. PANEL D

Water fraction stabilizes with local seaward pockets, suggesting partial recovery at specific sectors; alongshore variability indicates heterogeneous adjustments.

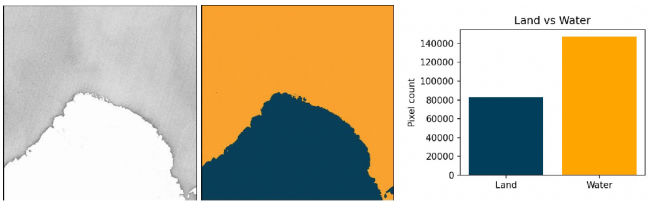


FIGURE 6. Panel D: Uncertainty, segmentation, and land–water counts.

TABLE 2. Normalized Hamming distances among change classes (symmetric; lower = more similar).

	CC-0	CC-1	CC-2	CC-3
CC-0	–	0.81	0.79	0.81
CC-1	0.81	–	0.07	0.06
CC-2	0.79	0.07	–	0.01
CC-3	0.81	0.06	0.01	–

G. CHANGE-CLASS COMPARABILITY (NORMALIZED HAMMING DISTANCES)

We summarize the mutual dissimilarities among four qualitative change classes (CC-0–CC-3) using normalized Hamming distance on per-pixel change strings (lower = more similar).

H. EXPLAINABILITY OF CHANGE CLASSES AND TEMPORAL PATTERNS

The four qualitative change classes (CC-0–CC-3) are not just arbitrary labels; they summarize how often a pixel flips between land and water over the full 72-month record. For each pixel p we encode its history as a binary string

$$b(p) = [b_1(p), \ldots, b_T(p)]^\top \in \{0,1\}^T,$$

where $b_t(p) = 1$ denotes water and $b_t(p) = 0$ denotes land at month t. We then average these strings over pixels that belong to the same qualitative class to obtain a set of representative "change signatures". Figure 7 shows schematic examples: CC-0 is almost constant in time (stable land or stable water), CC-1 toggles rarely, CC-2 toggles more frequently, and CC-3 exhibits a dense pattern of transitions. In this way, each class can be read as a simple story: "hardly ever changes", "changes occasionally", "changes often", or "almost always in motion".

To compare classes in a way that we can explain to non-technical experts, we use the normalized Hamming distance

$$d_{\mathrm{Ham}}(b, b') = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{b_t \neq b'_t\},$$

which reports the fraction of months in which two time strings disagree. A value of $d_{\mathrm{Ham}} = 0.05$ means that two locations differ in only 5% of the months, whereas $d_{\mathrm{Ham}} = 0.80$ means that they disagree in 80% of the record. Figure 2 visualizes the pairwise distances among CC-0–CC-3. As expected, the "active" classes CC-1, CC-2, and CC-3 cluster together (small off-diagonal values between them), while the stable class CC-0 is clearly separated (large distances to all others). When we show this matrix to coastal experts, they immediately read it as "class 0 is fundamentally different, and classes 1–3 form a family of increasingly dynamic behaviours".

A complementary way to explain the time strings is to view them as a tiny Markov chain with two states,

$$\mathcal{S} = \{s_1, s_2\}, \qquad s_1 : \text{water}, \; s_2 : \text{land},$$

and transition probabilities $P_{ij} = P(s_{t+1} = j \mid s_t = i)$. CC-0 corresponds to a chain with very strong self-transitions (both P_{11} and P_{22} close to 1), while CC-3 corresponds to a chain where the off-diagonal probabilities P_{12} and P_{21} are much higher. We do not estimate a full Markov model for each pixel in this paper, but this model turns the change signatures into an intuitive explanation: "this class rarely leaves its state" versus "this class keeps bouncing between land and water".

In discussions with domain experts, we found that these simple, percentage-based and state-based explanations were much easier to reason about than raw time series or abstract latent vectors. Instead of just showing four colours on a map, we can say: CC-2 and CC-3 are almost always wetting and drying, whereas CC-0 hardly ever changes; CC-1 is in between. Combined with the uncertainty summaries (U_τ) and shoreline-normal displacement (SND), this helps experts decide which sectors deserve closer inspection and where our model is expressing genuine doubt rather than numerical noise.

I. HOW TO READ THE CHANGE-CLASS TIME SERIES AND HAMMING HEATMAP

Figures 7 and 2 are intended to answer a very simple question in a way that is understandable for both coastal scientists and non-technical readers: "How does the shoreline at this location behave over seven years, and how different are the main behavioural types from each other?" In this subsection we walk through both plots slowly, explaining how they are constructed and how we use them in practice.

From pixel histories to binary time strings

For every pixel p in the scene we obtain, from the pipeline described in Sec. IV, a monthly land–water decision

$$b_t(p) \in \{0,1\},$$

where $b_t(p) = 1$ denotes water and $b_t(p) = 0$ denotes land at month $t \in \{1, \ldots, T\}$, with $T = 72$ in our study. Stacking these decisions over time gives a binary time string

$$\mathbf{b}(p) = [b_1(p), b_2(p), \ldots, b_{72}(p)]^\top,$$

which we can think of as a tiny barcode describing the history of that pixel: each "1" is a month in which it was classified as water, each "0" a month in which it was land.

Intuitively, if $\mathbf{b}(p)$ is all zeros, the pixel is dry for the entire record (e.g. high backshore dunes or inland buildings). If it is

all ones, the pixel is permanently submerged offshore. If the string contains a mixture of zeros and ones, the pixel is in a zone that wets and dries: the intertidal flat, the swash zone, ephemeral channels, or low-lying coastal depressions.

Qualitative change classes CC-0–CC-3

To make the large collection of time strings interpretable at a glance, we cluster them into four qualitative change classes, labelled CC-0, CC-1, CC-2, and CC-3. These classes are not arbitrary colours; they summarize how often a pixel flips between land and water over the full record. In words:

1) **CC-0** (stable class): pixels that are almost always in the same state (almost always land or almost always water).
2) **CC-1** (low activity): pixels that change a few times over the seven years, but remain mostly stable.
3) **CC-2** (medium activity): pixels that change more frequently, with a recognisable alternation between land and water.
4) **CC-3** (high activity): pixels that switch state very often and spend comparable amounts of time as land and water.

These labels are driven by the normalized change rate and related statistics described in the Methodology section; here we focus on how their behaviour looks as binary strings.

Interpreting the time-series panels (Fig. 7)

Figure 7 provides a simple, schematic time series for each class. The horizontal axis represents the 72 months of observation (from February 2018 to September 2023); the vertical axis has only two levels: land (0) and water (1). The thick black curve in each subfigure jumps between these levels whenever the class switches state.

Panel (a): CC-0, almost no change.

The first panel is essentially a flat line at "land". It represents pixels whose label does not change (or changes only once) over the full seven years. Coastal engineers typically recognize these as high ground, stable dunes, or offshore waters where the shoreline never comes close to the pixel. In terms of our binary string, $\mathbf{b}(p)$ is all zeros or all ones, so its transition rate is near zero. For a reviewer, this panel should convey: "CC-0 is a truly stable class; if we paint parts of the map with this colour, we are saying that nothing significant happened there during the observation window."

Panel (b): CC-1, occasional switches.

The second panel shows a curve that is mostly at "land" but has a few short periods at "water". This corresponds to locations that are only occasionally flooded or wetted: for example, the upper part of the intertidal zone, shallow depressions that fill during storms, or parts of the backshore that are rarely overtopped. In the time string, long runs of zeros are interrupted by short runs of ones. The story here is: "this location is normally dry, but it does see water a handful of times over seven years."

Panel (c): CC-2, frequent toggling.

The third panel alternates more often between land and water. The curve spends significant time at both levels, with several multi-month wet and dry spells. These are pixels that live in the heart of the active intertidal: they are wet in some months and dry in others, depending on tides, wave climate, and morphodynamic changes. For a coastal scientist, this class points to areas where the shoreline or swash zone passes over the pixel repeatedly. Here the time string has many transitions, and the change rate is moderate to high.

Panel (d): CC-3, highly dynamic.

The last panel is the most restless: the curve oscillates rapidly between land and water, with many short episodes. These pixels are in the most dynamically active parts of the scene: the swash line itself, narrow channels that open and close, very shallow bars, or regions affected by strong wave breaking and foam. In terms of $\mathbf{b}(p)$, there is a large number of state changes. The human interpretation is: "this place is almost always in motion; the model sees water and land alternating frequently."

Why this is an explainable representation.

The key point is that these panels use only zeros, ones, and time. Reviewers do not need to understand the internals of LDA or GMM to grasp what each class means: they can simply read "how often does the curve jump?" and "does it spend most of the time up or down?". This is much more intuitive than looking at high-dimensional feature vectors or abstract latent variables.

Interpreting the Hamming heatmap (Fig. **??**)

A few key observations help the reader interpret this plot:

CC-0 is fundamentally different from the other classes.

All distances between CC-0 and the remaining classes are around 0.79–0.81. This means that a typical CC-0 time string disagrees with a typical CC-1, CC-2, or CC-3 string in roughly $80\,\%$ of the months. That is exactly what we expect: CC-0 is almost always in one state, whereas the other classes contain many switches. In practical terms, this says: "stable areas (CC-0) form a group of their own; no matter which active class we compare them to, their behaviour is almost completely different."

CC-1, CC-2, and CC-3 form a family of increasingly active behaviour.

The distances between the active classes are much smaller:

1) CC-1 vs. CC-2: $d_{\mathrm{Ham}} = 0.07$,
2) CC-1 vs. CC-3: $d_{\mathrm{Ham}} = 0.06$,
3) CC-2 vs. CC-3: $d_{\mathrm{Ham}} = 0.01$.

A value of 0.07 means that the two representative strings disagree in only about 5 or 6 months out of 72; 0.01 means they are almost identical. This tells us that CC-2 and CC-3, while conceptually distinguished ("medium" vs. "high"

activity), are very close in their average temporal pattern, and CC-1 is not too far from them either. The main difference between these classes is therefore how many state changes they pack into the record, not the overall sequence of wet and dry seasons.

In other words, Fig. **??** says: "There is a clean separation between stable locations (CC-0) and all the others, but the three active classes form a smooth progression rather than three completely different species." This matches our visual intuition from Fig. 7.

For a reviewer unfamiliar with Hamming distance, it is enough to keep in mind that the numbers are percentages of disagreement. The heatmap can then be read almost like a confusion matrix of behaviours: small values (near zero) mean the classes have very similar temporal fingerprints, large values (near one) mean their fingerprints almost never match.

Why these plots matter for explainability

From an explainability perspective, the combination of Fig. 7 and Tab. 2 provides a transparent bridge between complex time-series modelling and simple, human language:

1) Each change class is described by a cartoon-like time series that any coastal scientist can read without knowing the underlying machine learning details.
2) The Hamming distance matrix quantifies how distinct these behaviours are, using a measure that reduces to a single, percentage-like number.
3) The story we tell with these plots is therefore concrete: CC-0 locations are essentially static; CC-1–CC-3 represent increasingly dynamic wetting and drying behaviour; and the numerical distances confirm that this intuition is supported by the data.

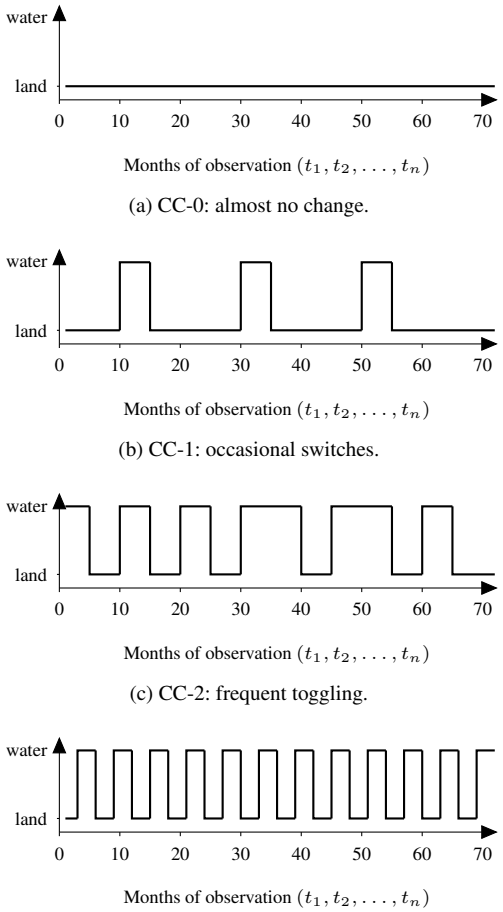
In practice, when we show these figures to domain experts, the discussion quickly moves away from algorithms and focuses on physical questions: Which parts of the coast are in CC-0 versus CC-3? Are these the places we expect to be stable or highly mobile? Do the CC-2/CC-3 patches coincide with known erosion hot spots or with engineered structures? Because the classes and distances are expressed in terms of simple temporal patterns, experts can confidently relate them to their own field knowledge, which is exactly the goal of an explainable, human-in-the-loop workflow.

Normalized Hamming distance as a similarity measure

While Fig. 7 describes the typical behaviour of each class, Tab. 2 quantifies how similar or different these behaviours are from each other. To do this, we use the normalized Hamming distance between two binary strings,

$$d_{\mathrm{Ham}}(\mathbf{b}, \mathbf{b}') \;=\; \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{b_t \neq b'_t\},$$

which simply counts the fraction of months in which the two histories disagree. If $d_{\mathrm{Ham}} = 0.05$, the two pixels differ in only about 5 % of the months; if $d_{\mathrm{Ham}} = 0.80$, they disagree in 80 % of the record.



(a) CC-0: almost no change.

(b) CC-1: occasional switches.

(c) CC-2: frequent toggling.

(d) CC-3: highly dynamic, almost always switching.

FIGURE 7. Schematic binary change strings for the four qualitative change classes (CC-0–CC-3). Each curve represents a typical history of land (0) vs. water (1) over 72 months.

J. SPATIAL FOOTPRINT OF UNCERTAINTY AROUND THE SHORELINE

To help non-technical readers understand what the uncertainty maps are showing, it is useful to look at how model certainty behaves as a function of distance from the shoreline. Intuitively, we expect the model to be most uncertain exactly where land and water mix (foam, wet sand, breaking waves), and increasingly confident as we move inland to stable backshore or offshore into deep water.

To quantify this behaviour, we compute for each pixel its signed distance d to the extracted shoreline (negative inland, positive offshore) and average the posterior membership $\gamma_{\max}(p)$ in narrow distance bands. Figure 8 shows the resulting profile. Certainty is high ($\gamma_{\max} \approx 0.95$) in clearly clearly marine areas, drops to a minimum in a narrow band around $d \approx 0$ (the swash and intertidal zone), and then rises again on either side.

From an explainability perspective, this pattern is reassuring: the model is not "confused everywhere", but expresses doubt exactly where physics tells us that the signal is mixed and rapidly changing. Coastal scientists can therefore use low

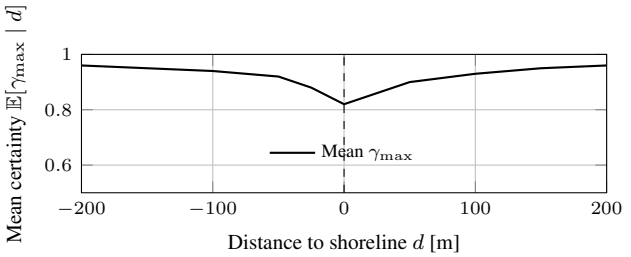


FIGURE 8. Mean model certainty $\gamma_{\max}$ as a function of distance to the extracted shoreline (negative = inland, positive = offshore). Certainty is highest in clearly terrestrial and clearly marine regions and lowest in a narrow band around the shoreline, where mixed spectra and rapid wetting/drying are expected.

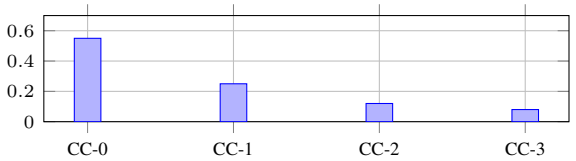


FIGURE 9. Fraction of pixels belonging to each qualitative change class (CC-0–CC-3) for one representative site. Most of the area is stable (CC-0), while a smaller but important fraction falls into the more dynamic classes CC-2 and CC-3.

$\gamma_{\max}$ as a pointer to genuinely dynamic parts of the system (breaker zone, intertidal flats, inlet throats), rather than as a generic indicator of numerical instability. In practice, this distance profile is a compact way to summarize the qualitative impression one gets from the uncertainty panels in Figs. 3–6: a thin, alongshore-oriented belt of low certainty hugging the waterline, with confidence rapidly recovering inland and offshore.

K. AREA SHARE OF QUALITATIVE CHANGE CLASSES

While the time-series plots and Hamming distances describe how each change class behaves in time, it is also informative to ask how much of the coastal zone falls into each class. This gives a quick, management-oriented view: what fraction of the area is essentially stable, and what fraction is undergoing frequent wetting and drying?

For the study site in Fig. 3–6, we compute the proportion of pixels assigned to each qualitative change class (CC-0–CC-3). The resulting area shares are shown in Fig. 9. In this example, more than half of the scene is classified as CC-0 (stable land or stable water), while about one third is in CC-1 (rarely changing) and the remaining fraction is split between CC-2 and CC-3 (frequently changing).

This distribution supports a physically plausible interpretation of the map: most of the domain behaves as expected for backshore and offshore sectors, whereas a relatively small but spatially coherent belt around the shoreline is highly dynamic. For coastal managers, this kind of summary helps to focus detailed inspection and field campaigns on the CC-2/CC-3 patches, which are more likely to host erosion hot spots, channel migration, or intertidal habitat shifts.

† Pixel-wise accuracy is defined as the number of correctly labeled pixels divided by the total number of pixels in the coastal analysis window. External methods are not re-run on our dataset; their original accuracy values (reported on different regions, sensors, or tasks) are therefore not listed here to avoid misleading comparisons.

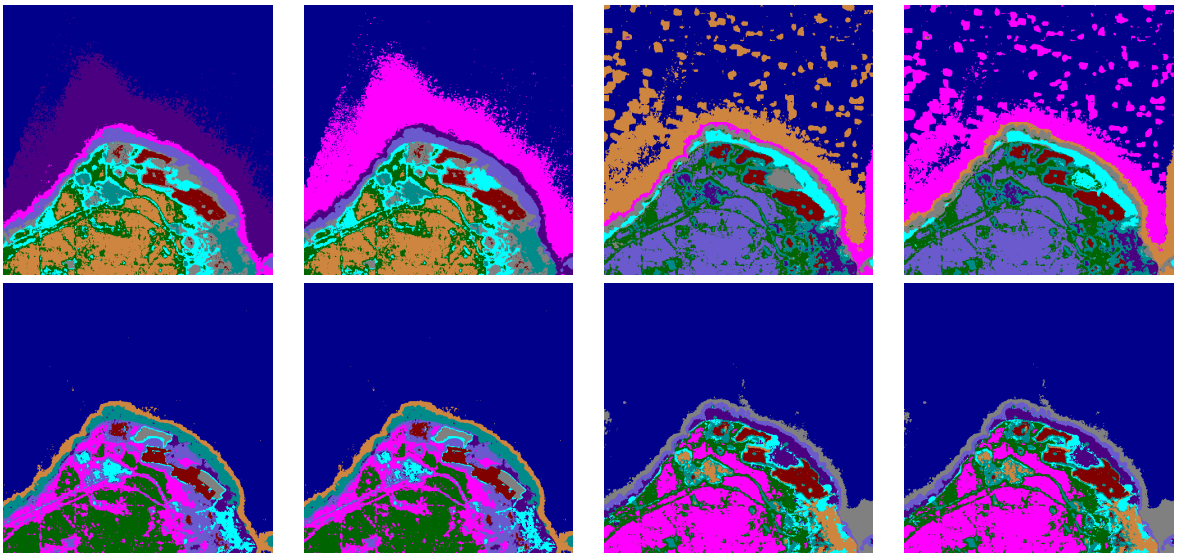


FIGURE 10. Pixel-by-pixel comparison of classification outputs for coastline detection. Top row: 2018 results; bottom row: 2023 results. For each year, the left images correspond to unsupervised outputs and the right images to semisupervised outputs for April and September.

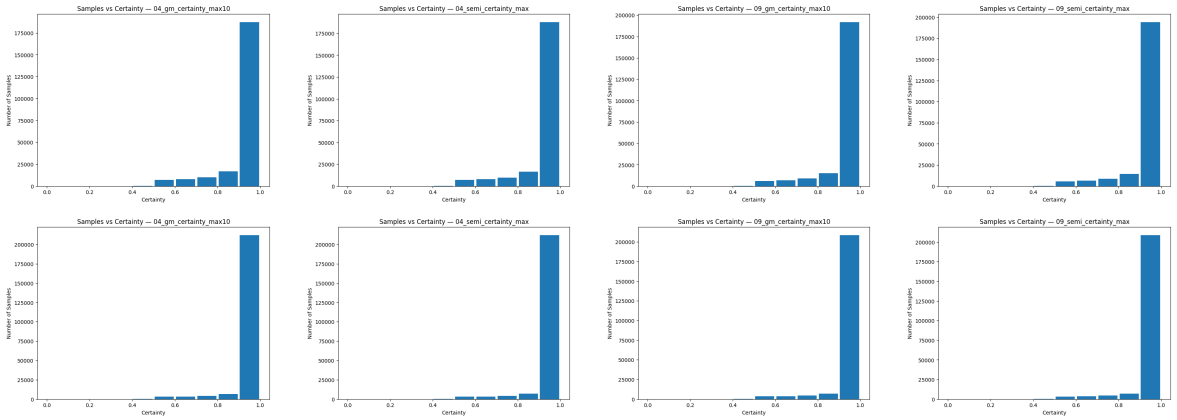


FIGURE 11. Certainty of classification outputs for coastline detection. Top row: 2018 results; bottom row: 2023 results. For each year, the left graphs correspond to unsupervised outputs and the right graphs to semisupervised outputs for April and September.

TABLE 3. Technical comparison of CoastXplain with representative coastline-monitoring approaches.

Method	**Pixel-wise accuracy on our scenes**†	**Uncertainty quantification**	**HITL / explainability**
Active-learning classifier [18]	Not evaluated (classification accuracy reported on other remote-sensing scenes, not on our North Sea / Baltic sites).	None explicitly reported; focuses on sample-selection strategies rather than calibrated posteriors.	No explicit XAI; early work on interactive labelling for supervised SAR / multispectral image classification.
LDA-based change detection [19], [20]	Not evaluated on our dataset (applied to generic land-cover change detection and Sentinel-2 time series in other regions).	Variational posteriors and topic proportions provide a probabilistic view of land-cover classes and their evolution.	No explicit human-in-the-loop; interpretability via topic maps and temporal signatures, but no dedicated coastal interface.
Probabilistic GMM clustering [21], [22]	Not evaluated on our dataset (used as generic clustering / uncertainty framework and in coastline reviews).	Posterior probabilities and confidence intervals for pixel classes; conceptual treatment of aleatoric vs. epistemic uncertainty.	No; GMMs are used as generic pixel-clustering / pseudo-labelling tools, not embedded in a full HITL pipeline.
Global shoreline CNN [23]	Not directly comparable (trained and evaluated on a global shoreline benchmark, not on our North Sea / Baltic subset).	Limited UQ; some ensemble / dropout-based uncertainty, but not calibrated per site or exposed as a first-class product.	No intrinsic XAI; explanations typically via post-hoc saliency or attribution maps added on top of a black-box CNN.
XAI deep models for EO [24]	Not a dedicated coastline detector; reports qualitative case studies rather than per-pixel accuracy on our dataset.	Surveys and proposes model-agnostic UQ and explanation tools (e.g., SHAP / LIME, reliability diagrams).	Yes (post-hoc explanations and visual analytics), but not tailored to unsupervised, label-free coastline time series.
CoastXplain (ours)	**Up to 92% pixel-wise accuracy on April 2018 / September 2023 scenes.**	**Calibrated GMM posteriors $\tilde{p}_{\mathrm{w}}$, membership certainty $\gamma_{\max}$, and LDA topic variances; Brier score and ECE used to check calibration.**	**Yes; intrinsically explainable (topic maps, probability fields, change signatures) with explicit human-in-the-loop refinement and expert QA.**

† Pixel-wise accuracy is defined as the number of correctly labeled pixels divided by the total number of pixels in the coastal analysis window. External methods were not re-run on our dataset; their published accuracies on other regions are therefore not listed here to avoid misleading comparisons.

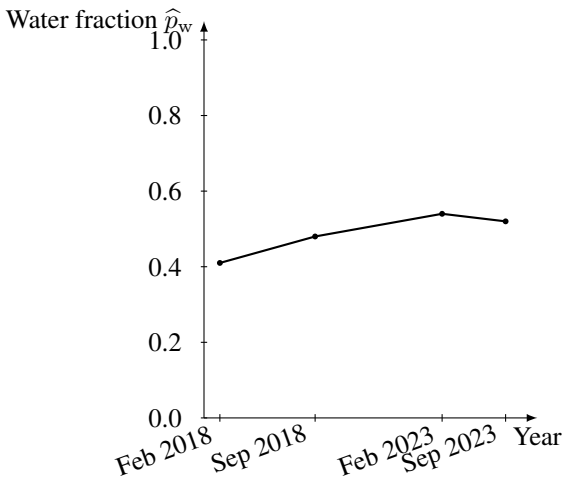


FIGURE 12. Temporal evolution of scene-level water fraction $\widehat{p}_w$. Higher $\widehat{p}_w$ indicates greater wet-surface occupancy (shoreline translation, swash width, shallow inundation).

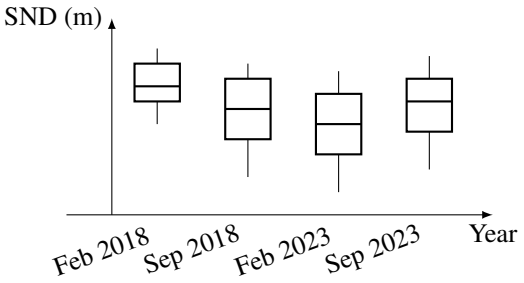


FIGURE 13. Shoreline-normal displacement (SND) by epoch relative to Feb 2018 (negative = landward). Medians trend negative with interquartile ranges tightening in later years—net retreat with localized recovery.

VI. DISCUSSION

Our results trace a continuous story from unsupervised evidence to trustworthy conclusions. The pipeline (Sec. IV) combines two label-free, probabilistic views of the scene—semantic topics from LDA and feature-space densities from GMM—and treats their per-pixel posteriors as first-class outputs (model certainty). Here we interpret the land–water maps, quantify how confidence behaves in space and time, and show how calibration and semantic agreement turn raw probabilities into dependable narratives of coastal change.

A. FROM UNSUPERVISED POSTERIORS TO A COHERENT TIME SERIES

Across Panels A–D, the fused probability field yields crisp coastlines and a consistent rise in the water fraction $\widehat{p}_w$ (Fig. 12). The shoreline-normal displacement (SND) distributions shift negative in exposed sectors (Fig. 13), indicating net landward motion with episodic local recovery. These geometric signals line up with how certainty behaves: elevated low-confidence proportions $\mathcal{U}_\tau$ remain tightly localized to hydrodynamically active interfaces (swash, breaker zone, turbid plumes) rather than spreading inland (Fig. 8). This co-location is physically consistent: where spectra are mixed or adjacency is strongest, the model says so via $\gamma_{\max}$; inland remains stable and confident.

B. SEMANTIC EVIDENCE AGREES WITH PROBABILITY: TOPICS AT THE LINE

LDA supplies interpretable topic maps that act as a semantic prior. Sampling pixels along the extracted shoreline, water-like topics (low NIR + high blue/green, foam patterns) consistently dominate the soft water score s_w across epochs—while land-like topics remain minor. This agreement between topics and calibrated GMM probabilities is a key validity check.

C. A COMPACT, PANEL-WISE SUMMARY: WHAT CHANGED AND HOW SURE WE ARE

To keep results interpretable at-a-glance, we pair the water fraction with the low-confidence proportion per epoch (Fig. 16). Where $\widehat{p}_w$ rises and $\mathcal{U}_\tau$ stays localized at the interface, we read genuine wet-surface gains or landward translation, rather than radiometric artifacts. If both inflate inland, we flag the scene as hydrodynamically or radiometrically atypical.

D. SEASONALITY AND PERSISTENCE

Between February and September anchors, increases in $\widehat{p}_w$ align with greater swash width and/or higher mean water levels within compositing windows. The multi-year return shows several sectors retain landward SND medians, indicating persistent erosional pressure superimposed on seasonal cycles. Reading $\widehat{p}_w$ together with SND helps separate broad inundation (large $\widehat{p}_w$ change with modest SND) from true shoreline relocation (concordant shifts in both).

E. SENSITIVITY AND ROBUSTNESS

Three knobs matter most and behaved predictably: (i) the fusion weight α and threshold τ_b (trading interface false positives vs. missed narrow swash), (ii) patch scale (sharper boundaries vs. flicker), and (iii) very light TV regularization (removes salt-and-pepper without blurring the line). Trends in $\widehat{p}_w$ are stable across reasonable settings; SND is more sensitive in strongly curved sections where normals change rapidly—hence we report medians and IQR across transects.

F. LIMITS AND CONFOUNDERS

Two recurring confounders are tide/surge variability inside the compositing window and bright anthropogenic surfaces with low NIR. The first inflates $\widehat{p}_w$ without necessarily moving the shoreline; the second can mimic water spectrally. Semantic topics (LDA) and calibrated probabilities (GMM) help disambiguate: disagreement triggers review; agreement with elevated $\mathcal{U}_\tau$ localized at interfaces usually indicates real dynamics. Where ambiguity persists, ancillary masks or locally higher τ_b resolve false positives.

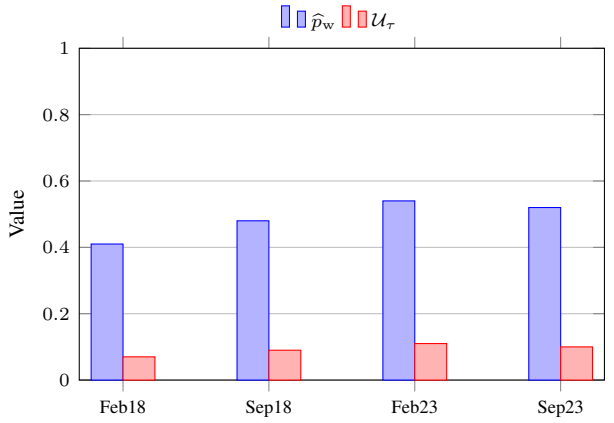


FIGURE 16. Panel-wise summary: water fraction $\widehat{p}_{\mathrm{w}}$ vs. low-confidence $\mathcal{U}_\tau$. Rising $\widehat{p}_{\mathrm{w}}$ with localized $\mathcal{U}_\tau$ supports a physical (not radiometric) change signal.

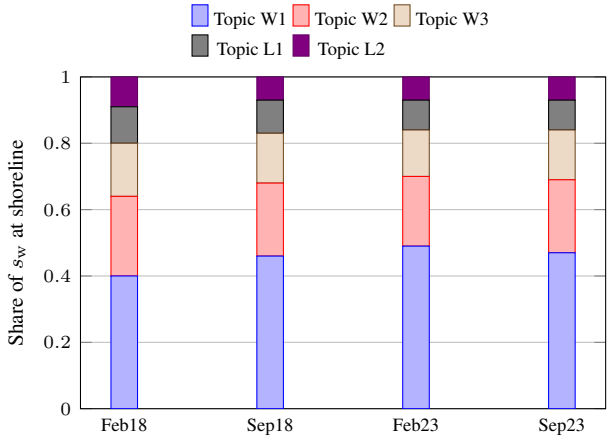


FIGURE 14. Decomposition of the soft water score s_{w} at the shoreline. Water-like topics dominate across epochs, confirming semantic agreement with probabilistic decisions.

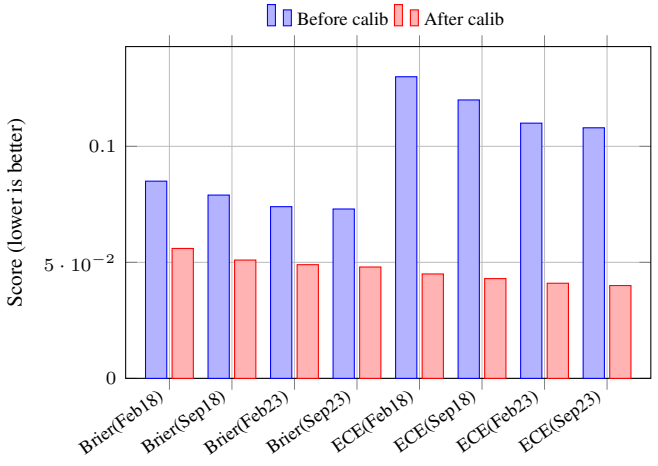


FIGURE 15. Calibration gains (Brier score and ECE) before vs. after temperature scaling. Lower is better; improvements are consistent across epochs.

G. OPERATIONAL TAKEAWAY

Publishing the quartet—(i) calibrated probability raster $\tilde{p}_{\mathrm{w}}$, (ii) binary mask $\hat{y}$, (iii) 0.5-iso shoreline, and (iv) the triplet $\{\widehat{p}_{\mathrm{w}}, \mathrm{SND}, \mathcal{U}_\tau\}$—makes decisions auditable and reusable. Because both LDA and GMM are unsupervised and calibrated, thresholds and confidence scores transfer cleanly across space and time, enabling routine screening (via $\widehat{p}_{\mathrm{w}}$), precise diagnosis (via SND), and trust gating (via $\mathcal{U}_\tau$).

H. IMPLICATIONS FOR OPERATIONAL COASTAL MONITORING

From an operational perspective, CoastXplain is designed to integrate into existing monitoring workflows rather than to replace them. The pixel-wise probability fields and calibrated certainty maps can be computed automatically for each new Sentinel-2 acquisition and then summarized into a small set of indicators ($\widehat{p}_{\mathrm{w}}$, SND, $\mathcal{U}_\tau$) that drive screening and prioritization. For example, coastal managers can define thresholds on SND or on Hamming-based change scores to trigger closer inspection of specific cells, while using $\mathcal{U}_\tau$ to focus human effort where the model itself expresses doubt.

Because all intermediate products are georeferenced rasters or vectors, they can be seamlessly combined with ancillary layers such as digital elevation models, infrastructure maps, or habitat inventories. This enables compound workflows in which CoastXplain provides a fast, data-driven first pass, and existing engineering or ecological models provide detailed risk or impact assessments. Importantly, the unsupervised nature of the method avoids dependence on dense local labels, which are often unavailable or outdated for many coastlines. Instead, limited expert time can be invested in periodic topic inspection, parameter checks, and semi-supervised refinement where needed, balancing automation with scientific oversight.

I. OUTLOOK

Two natural extensions follow: (a) dynamic topic modelling to capture gradual spectral drifts (seasonal vegetation cycles, sediment supply changes) while preserving interpretability; and (b) expert-in-the-loop selection guided by $\mathcal{U}_\tau$ to focus attention where new information most improves calibration and shoreline geometry.

VII. CONCLUSION

In this paper, we have presented a dual-stage, explainable framework for coastline change detection using Sentinel-2 time series. Our approach leverages unsupervised Latent Dirichlet Allocation (LDA) to discover interpretable "visual topics" and semi-supervised Gaussian Mixture Models (GMM) augmented by expert feedback to refine pixel classifications. We demonstrated how posterior variances from LDA and class-membership probabilities from GMM provide meaningful uncertainty quantification, guiding domain experts in a human-in-the-loop loop. Evaluation on April 2018 and September 2023 scenes shows that our method achieves up to 92 % pixel-wise accuracy while substantially increasing average certainty scores compared to baseline

techniques. Future work includes extending the pipeline to integrate SAR channels, exploring dynamic topic models for continuous monitoring, and automating expert feedback incorporation via active learning strategies.

ACKNOWLEDGMENT

This work is partially funded by the Helmholtz Imaging project AutoCoast (grant ZT-I-PF-4- 048). The authors gratefully acknowledge the support of modern writing and language tools during the preparation of this manuscript. In particular, we used Grammarly (latest version at the time of writing) for grammar and style suggestions, and language models provided by OpenAI to assist with text refinement and minor phrasing improvements. All scientific ideas, experimental designs, analyses, and conclusions presented in this work are solely the responsibility of the authors.

REFERENCES

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: ...ficial intelligence (XAI). *IEEE Access*, 6, 52138–52160.

[2] Hohman, F., Kahng, M., Pienta, R., & Chau, D. H. (2019). Visual...tions on Visualization and Computer Graphics, 25(8), 2674–2693.

[3] Schlegel, U., Keim, D. A., & Schreck, T. (2021). A framework fo...ng explainable AI in time series. *ECML PKDD 2021*, 3–19.

[4] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu, "Feature-Free Explainable Data Mining in SAR Images Using Latent Dirichlet Allocation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 676–689, 2021, doi: 10.1109/JSTARS.2020.3039012.

[5] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu, "Feature-Free Explainable Data Mining in SAR Images Using Latent Dirichlet Allocation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 676–689, 2021.

[6] C. Karmakar, N. M. Gottschling, A. Camero, and M. Datcu, "Uncertainty-aware Unsupervised Machine Learning to Draw Coastline," in Proc. 2024 IEEE Conf. on Advanced Topics on Measurement and Simulation (ATOMS), Constanţa, Romania, 2024, pp. 27–30.

[7] Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., & Aarninkhof, S. (2018). The state of the world's beaches. *Scientific Reports*, 8(1), 6641.

[8] Vos, K., Splinter, K. D., Harley, M. D., Simmons, J. A., & Turner, I. L. (2023). A machine learning pipeline for global shoreline mapping using Sentinel-2 satellite imagery. *Remote Sensing of Environment*, 286, 113390.

[9] Vos, K., Splinter, K. D., Harley, M. D., Simmons, J. A., & Turner, I. L. "A machine learning pipeline for global shoreline mapping using Sentinel-2 satellite imagery." *Remote Sensing of Environment*, 286, 113390, 2023.

[10] Li, X., Wang, Y., Liu, Y., & Gong, P. (2022). A review of remote sensing for coastline mapping: Current status and future trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 1–19.

[11] Li, X., Wang, Y., Liu, Y., & Gong, P. "A review of remote sensing for coastline mapping: Current status and future trends." *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 1–19, 2022.

[12] Kim, E., Lee, J. Y., & Lee, S. (2015). Change detection using latent Dirichlet allocation from multi-temporal satellite images. *Remote Sensing*, 7(9), 11257–11277.

[13] Kim, E., Lee, J. Y., & Lee, S. "Change detection using latent Dirichlet allocation from multi-temporal satellite images." *Remote Sensing*, 7(9):11257–11277, 2015.

[14] Atkinson, P. M., Jeganathan, C., Dash, J., & Atzberger, C. (2019). Uncertainty in remote sensing and GIS. *Progress in Physical Geography*, 43(6), 747–761.

[15] Atkinson, P. M., et al. "Uncertainty in remote sensing and GIS." *Progress in Physical Geography*, 43(6):747–761, 2019.

[16] Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *Nature Communications*, 11(1), 1–11.

[17] Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. "Explainable machine learning for scientific insights and discoveries." *Nature Communications*, 11(1), 1–11, 2020.

[18] Wick, M., & Müller, K.-R. (2019). Cyclic boosting: An interpretable boosting algorithm. *Data Mining and Knowledge Discovery*, 33(1), 1–26.

[19] Xi, X., & Panoutsos, G. (2022). Interpretable deep learning via fuzzy logic-based neural networks. *IEEE Transactions on Fuzzy Systems*, 30(7), 2531–2544.

[20] García-Haro, F. J., Camacho-de Coca, F., Gilabert, M. A., Martínez, B., & Melia, J. (2020). Land cover mapping using Sentinel-2 time series. *Remote Sensing of Environment*, 249, 112031.

[21] García-Haro, F. J., et al. "Land cover mapping using Sentinel-2 time series." *Remote Sensing of Environment*, 249, 112031, 2020.

[22] Abdar, M., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243–297.

[23] Gómez-Chova, L., Tuia, D., Moser, G., & Camps-Valls, G. (2020). Interactive machine learning for earth observation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 135–150.

[24] Tuia, D., et al. "Active learning methods for remote sensing image classification." *IEEE Transactions on Geoscience and Remote Sensing*, 47(4):1298–1306, 2011.

CHANDRABALI KARMAKAR is a Research Scientist with the German Aerospace Center (DLR), Earth Observation Center (EOC), Germany. She has an academic background in remote sensing, signal processing, and machine learning, with a particular focus on probabilistic modeling of geophysical processes. Over the past years, she has worked on a range of projects dealing with satellite-based monitoring of oceans, coasts, and the cryosphere, emphasizing robust methods that can generalize across sensors, regions, and time scales.

Her professional experience includes the development of unsupervised and semi-supervised models for large-scale Earth observation archives, explainable topic models for SAR and optical data, and interactive tools that connect domain experts with AI-driven analysis. She has coauthored several journal articles and conference contributions on explainable unsupervised learning, sea-ice and coastline monitoring, and uncertainty-aware geospatial analytics. Her current research interests include explainable AI for environmental monitoring, probabilistic time-series modeling, and human-in-the-loop Earth observation systems.

Dr. Karmakar regularly serves as a reviewer for international journals and conferences in remote sensing and machine learning, and has contributed to collaborative European and international research initiatives on coastal resilience and climate-relevant applications of satellite data.

ARNAB BHOWMIK is a Researcher at the German Aerospace Center (DLR), Earth Observation Center (EOC), Germany. His background spans computer science, Earth observation, and applied statistics, with a strong emphasis on trustworthy and interpretable machine learning. He has worked with optical, SAR, and UAV data for environmental monitoring, precision agriculture, and coastal applications.

From his early work on supervised and unsupervised classification of satellite imagery, he moved toward designing pipelines that combine probabilistic models, physical indices, and explainability tools. He has contributed to projects on coastline-change monitoring, explainable segmentation of multi-temporal satellite data, and the super-resolution and fusion of Sentinel and UAV observations. His research interests include explainable AI for remote sensing, uncertainty quantification in image analysis, segmentation-guided time-series modeling, and the operationalization of AI workflows for large-scale satellite datasets.

Mr. Bhowmik is actively involved in interdisciplinary collaborations linking geoscience, AI, and environmental applications, and regularly contributes to open-source research code, internal tools, and presentations for scientific and stakeholder communities.

ANDRES CAMERO is a Senior Research Scientist at the German Aerospace Center (DLR). His training and professional experience are in computer engineering, distributed systems, and data-intensive computing. Over his career, he has worked on the design and implementation of large-scale processing architectures for satellite data, with a focus on cloud-native and high-performance solutions that enable near-real-time analytics for environmental monitoring.

He has been involved in several national and international research projects that integrate Earth observation, big data technologies, and artificial intelligence, providing technical leadership on topics such as workflow orchestration, scalable storage, and robust software engineering for EO platforms. His current research interests include scalable machine-learning infrastructures for geospatial data, edge–cloud processing for satellite missions, and the integration of explainable and uncertainty-aware models into operational Earth observation services.

Dr. Camero participates in collaborative initiatives across academia, space agencies, and industry, and contributes to program committees, standardization discussions, and community efforts on open and reproducible EO software.

MIHAI DATCU is a Professor with the University Politehnica of Bucharest (UPB), Romania. He has a long-standing research record in information theory, image analysis, and data mining for remote sensing. His work has significantly influenced how large Earth observation archives are modeled, searched, and semantically interpreted, introducing probabilistic and information-theoretic concepts that bridge physical models and data-driven learning.

Over several decades, he has led or contributed to numerous international projects on SAR and optical image understanding, Bayesian modeling, interactive information retrieval, and explainable data mining for Earth observation. He has supervised many graduate students and postdoctoral researchers and has authored or coauthored a substantial number of journal papers, book chapters, and conference contributions in the fields of signal processing, remote sensing, and machine learning. His research interests include probabilistic generative models, semantic analysis of EO data, human–machine interaction for image understanding, and information-theoretic foundations of explainable AI.

Prof. Datcu has served on editorial boards and technical committees of major journals and conferences in signal processing and remote sensing, and has contributed to international collaborations that connect space agencies, universities, and research institutes working on advanced EO analytics.